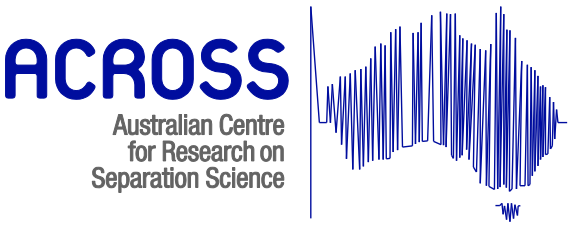

Free-solution Capillary Electrophoresis for the Structural Characterization of Proteins

Presented by:

Mar-Dean Du Plessis

BSc (Science - Biochemistry and Molecular Biology)

A thesis submitted in fulfillment of the requirement of

Master of Research

Principal Supervisor: Patrice Castignolles

Co-Supervisors: Marion Gaborieau, Christopher Jones

Western Sydney University, Australia

December 2018

Aan Ouma Lien en Oupa Frans. Dankie vir alles. Lief en mis julle oneindiglik.

Acknowledgements

Firstly, I would like to thank my supervisors for their guidance over the course of this degree. To Patrice and Marion, thank you for your endless patience and continuous encouragement as I attempted to tackle the numerous concepts that were far beyond my level of understanding. Thank you for always making time for discussions, both work and non-work related. The level of care and commitment you displayed for the team is what made me want to be a part of it in the first place- oh, and a little bit of the chemistry too! I am so grateful for all the opportunities the two of you have given me. To Chris, thank you so much for the time you spent, despite having a very busy schedule, offering your insights, suggestions and feedback for the improvement of this project.

To Dr Asa Jamting, there is a lot to be acknowledged when it comes to you. Firstly, thank you and Dr Victoria Coleman for the opportunity to have interned at NMI. Although it was short and the work I did on nanoparticles extended far beyond my comfort zone, having you as a mentor allowed me to grow on an intellectual and professional level. I would also like to acknowledge Professor Thomas Millar for all the insightful discussions and taking the time to teach me what it means to be passionate about research and to show it in one's work.

Thank you to the Macromolecular Characterisation Team and the rest of the members in the SSH postgrad office for all the interesting conversations and creating a supportive work environment. To James Oliver, thank you for all your help, especially in the final stages of writing up. I would also like to thank all the technical staff at WSU for all their support and the shared hand-gestures as we passed each other in common areas. Additionally, I would like to acknowledge Western Sydney University for the research funding and scholarship.

To my dearest Tennille, thank you for always having time for me, whether it's to listen to me complain or just to go get yet another schnitty! Your endless support and constant supply of smarties, mnm's and milo has honestly helped me through some of the most challenging times. I don't know how I would've gotten through these past few months without you!

To the insensitive members of Party Car, Bollie (Charline), Gramps (Joel), Shags (Matt) and CP (Lola), you all suck for the insane amount of sleep you deprived me of during the crucial start of this degree! Regardless, our times together were some of the most memorable ones yet. Thank you for forcing me to take some time to live a little. RIP to the Batmobile.

To BIG MAC (Aidaen), it has been an honor working with you for these 2 long years. I have only one question for you, where is the MANUSCRIPT?! Thank you for all the support and advice! Goodluck with finishing your thesis and have fun being a dad! To Joel (aka Robin), I know you already like to take ALL the credit for MY achievements, so don't let what I'm about to say make your big head even bigger! Simply put, thank you so much for everything. Not only did you persuade me to do this degree, you have stood by my side throughout the entire thing, as a friend and a mentor. Thank you for believing in me and leading me to do so myself. I'm honored to have been known as your "mini-me".

To my family, Tania, Ockert and Buddy. I'm sorry for never being home! I know you all must miss me incredibly (especially you, Buddy)! But don't worry, I will be home more now, so we can hang out ALL THE TIME! Thank you for your endless support and patience as I pursued my dreams to become the smartest person in the house! Without you, I would never have been able to do any of this. You listened to me ramble on about work even when it made no sense to you and offered advice when I needed it most. There are no words to describe my gratitude and love for the three of you.

Statement of Authentication

The work presented in this thesis is, to the best of my knowledge and belief, original except as acknowledged in the text. I hereby declare that I have not submitted this material, either in full or in part, for a degree at this or any other institution.

……  ……

Mar-Dean Du Plessis

December 2018

Research Output

Peer-reviewed Journal Articles in Preparation:

A.B. Grosas, **M. Du Plessis**, J.J. Thevarajah, M. Gaborieau, J. Carver & P. Castignolles, 'Characterization of Protein Structures from the Distribution of Electrophoretic Mobilities: A Comparative Study with the Distributions of Molar Masses'

J.J. Thevarajah, M.P. Van Leeuwen, **M. Du Plessis**, S. Harrisson, M. Gaborieau, & P. Castignolles, *Review*, 'Analytical Techniques for the Measurement of Polymer Dispersity'

Posters presentations as presenter:

M. Du Plessis, A.B. Grosas, J.J. Thevarajah, J.D. Oliver, Å. Jämting, P. Castignolles & M. Gaborieau, Corroborating Capillary Electrophoresis with Field Flow Fractionation for the Structural Characterisation of Proteins, Field Flow Fractionation of Polymers, Proteins and Nanoparticles Workshop, Western Sydney University, Parramatta (Australia), June 2018

M. Du Plessis, A.B. Grosas, J.J. Thevarajah, J. Carver, P. Castignolles & M. Gaborieau, Determination of the Conformational Heterogeneity of Proteins by Free-Solution Capillary Electrophoresis, Royal Australian Chemical Institute (RACI) NSW One-day Symposium on Analytical & Environmental Chemistry, CSIRO, Lindfield (Australia), April 2018

M. Du Plessis, A.B. Grosas, J.J. Thevarajah, J. Carver, P. Castignolles & M. Gaborieau, Structural Characterisation of Proteins using Free Solution Capillary Electrophoresis, Sydney Surfaces and Soft Stuff Meeting (SASSY), University of New South Wales, Kensington (Australia), June 2017

M. Du Plessis, A.B. Grosas, J.J. Thevarajah, J. Carver, P. Castignolles & M. Gaborieau, Structural Characterisation of Proteins using Free Solution Capillary

Electrophoresis, 24th Annual Royal Australian Chemical Institute (RACI) Analytical & Environmental Division R&D Topics Conference, Western Sydney University, Parramatta (Australia), December 2016

Posters presented by others:

M. Du Plessis, A.B. Grosas, J.J. Thevarajah, P. Castignolles & M. Gaborieau, Adsorption of Proteins onto Silica Surface during Characterisation using Capillary Electrophoresis, Sydney Surfaces and Soft Stuff Meeting (SASSY), University of New South Wales, Kensington (Australia), June 2018

Awards and Professional Memberships

1.1. Awards

- 2018 Runner up "People's choice" for best poster pitch: Field Flow Fractionation of Polymers, Proteins and Nanoparticles Workshop, Western Sydney University, Parramatta (Australia), June 2018
- 2016 Second place prize for Best Poster Presentation: 24th Annual Royal Australian Chemical Institute (RACI) Analytical & Environmental Division R&D Topics Conference, Western Sydney University, Parramatta (Australia), December 2016

1.2. Professional Memberships

- Royal Australian Chemical Institute (RACI)- Postgraduate student member
- Australian Nanotechnology Network (ANN)
- Golden Key International Honours Society
- Australian Centre for Research on Separation Science (ACROSS)
- Western Sydney University Medical Sciences Research Group (MSRG)

Table of Contents

List of Tables

List of Figures

Abbreviations

Abbreviation	Definition
α-LA	α-Lactalbumin
$Đ$	Dispersity of molar mass distribution
μ	Electrophoretic mobility
μ_w	Weight-average electrophoretic mobility
σ	Standard deviation
ADH	Alcohol dehydrogenase
AF4	Asymmetric flow field flow fractionation
BGE	Background electrolyte
BSA	Bovine serum albumin
CBB	Coomassie Brilliant Blue
CD	Circular dichroism
CE	Free-solution capillary electrophoresis
CE-CC	Free-solution capillary electrophoresis in the critical conditions
CN-PAGE	Clear-native polyacrylamide gel electrophoresis
$D(W(\mu))$	Dispersity of electrophoretic mobility distributions

D	Diffusion coefficient
DNAse I	Deoxyribonuclease I
*d*n/*d*c	Refractive index increment
EOF	Electroosmotic flow
FFF	Field flow fractionation
FWHM	Full width at half peak height (maxima)
HF5	Hollow-fiber flow field flow fractionation
LC-CC	Liquid chromatography in the critical conditions
MALS	Multi-angle light scattering
M_n	Number-average molar mass
MS	Mass spectrometry
M_w	Weight-average molar mass
NAD^+	Nicotinamide adenine dinucleotide
NaPi	Chloride-containing sodium phosphate buffer
NMR	Nuclear magnetic resonance
PAA	Poly(acrylic acid)
PAGE	Polyacrylamide gel electrophoresis
PB	Sodium phosphate buffer

PDI	Polydispersity index
pI	Isoelectric point
PM	Pressure mobilization
R^2	Correlation coefficient
*R*f	Retention factor
RI	Refractive-index
RSD	Relative standard deviation
SD	Standard deviation
SDS	Sodium dodecyl sulfate
SEC	Size-exclusion chromatography
TDA	Taylor dispersion analysis
UV	Ultraviolet
V	Voltage
v	Velocity
$W(\mu)$	Weight distribution of electrophoretic mobility

Table of Contributions

Chapter	Contributors	Contributors and statement of contributions
CHAPTER 2: This chapter is being prepared for submission to a peer-reviewed journal.	Mar-Dean Du Plessis (MDP), Aidan B Grosas (ABG), Joel J Thevarajah (JJT), John Carver (JC), Marion Gaborieau (MG), Patrice Castignolles (PC) & Christopher Jones (CJ)	**Conceptualization and design:** The idea for this publication was conceptualized by CJ, MG & PC following an undergraduate research project. ABG, PC and MG were the primary contributors to the design. **Experimental execution and data analysis:** JJT and PC provided training and assistance in conducting CE experiments and data analysis. MDP performed 70% of the CE experiments and the associated data treatment. ABG performed 30% of CE experiments and all the SEC-MALS experiments, as well as the associated data treatment. **Result interpretation:** All the authors were involved in the discussion and overall interpretation of the results. **Drafting and production**: ABG wrote the first draft of the introduction and experimental sections. MDP contributed in writing of the introduction and the experimental section. MDP wrote the first draft of the results and discussion section as

		well as the conclusion. MDP, ABG, MG and PC were involved in the revision and refinement of a final draft.
Chapter 3	Mar-Dean Du Plessis (MDP), Marion Gaborieau (MG) & Patrice Castignolles (PC)	**Conceptualization and design:** MDP, MG & PC were involved in the conceptualization and design of the chapter. **Experimental execution and data analysis:** PC provided assistance with the experimental design as well as the data treatment. MDP performed all of the experiments and the associated data treatment. **Result interpretation:** All of the contributors were involved in the discussion and interpretation of the results. **Drafting and production**: MDP performed all the background research and wrote a complete first draft of the chapter. MG and PC revised and guided the work during the production of a final draft.
Chapter 4	Mar-Dean Du Plessis (MDP), James D Oliver (JDO), Åsa Jämting (AJ),	**Conceptualization and design:** MDP, CJ, MG & PC were the primary contributors involved in the conceptualization and design of the overall chapter. **Experimental execution and data analysis:**

	Christopher Jones (CJ), Marion Gaborieau (MG) & Patrice Castignolles (PC)	AJ and JDO provided training in conducting FFF experiments. MDP, JDO and AJ were all involved in conducting the AF4 experiments. MDP performed the associated data treatment with the guidance of PC and JDO. CJ provided assistance in conducting PAGE experiments and result analysis. MDP performed all PAGE, CE and PM experiments as well as the associated data treatment. **Result interpretation:** MD, PC, JDO and MG were involved in the discussion and interpretation of AF4 results. MD, PC, CJ and MG contributed to the interpretation and discussions concerning PAGE and CE results. MD, PC and MG were involved in the discussion of PM results. **Drafting and production**: MDP performed all the background research and wrote a complete first draft of the chapter. JDO, CJ, MG and PC revised and guided the work during the production of a final draft.

Abstract

Proteins constitute a broad class of macromolecules which are abundant in all cells and are responsible for a wide variety of biological processes. Exploring the molecular mechanisms of these biological processes is pivotal, not only for understanding protein-related diseases and disorders, but also for the advancement of the biotechnological applications of proteins.

Various techniques can be used for the structural characterization of proteins at both the atomic and molecular levels. Many proteins are structurally dynamic and thus may exhibit varying degrees of heterogeneity with regards to size, molar mass and molecular conformation. Currently, size-exclusion chromatography (SEC) coupled to multi-angle light scattering (MALS) is the most common technique used to study protein heterogeneity. Distributions of hydrodynamic volumes and molar masses can be obtained from SEC-MALS, however this technique faces several limitations such as high cost, poor reproducibility and adsorption, to name a few. Additionally, many proteins also undergo more subtle changes to their structures (such as domain rearrangements) which do not necessarily impact their overall size/molar mass.

This project was therefore aimed at establishing a new method using free-solution capillary electrophoresis (CE) for the structural characterization of proteins, to be used complementarily to SEC. CE is a robust, cost- and time-effective technique that requires small sample amounts requiring minimal preparation. Additionally, the mechanism of separation enables distributions of electrophoretic mobilities to be obtained. CE in the critical conditions (CE-CC) allows for separation independent of molar mass, offering insight into the heterogeneity of a protein with regards to molecular conformations.

Six globular proteins that have already been widely-characterized through various other techniques were selected for this project based on their isoelectric points and

the substantial amount of information already known about them. SEC-MALS of these proteins was carried out by collaborators from the Australian National University (ANU). The data from CE was used to obtain weight distributions of electrophoretic mobilities for selected proteins, and from this, their average electrophoretic mobilities were determined. The dispersion of the electrophoretic mobility distributions was quantified and expressed as dispersity values and standard deviations. Dispersity values were calculated using a calculation analogous to the dispersity of molar mass distribution M_w/M_n. The findings from these experiments reveal that some proteins may be relatively homogenous in terms of molar mass yet highly disperse in terms of electrophoretic mobility (and vice versa).

For the accurate characterization of proteins in solution, true solutions need to be obtained prior to analysis. Proteins are often converted to lyophilized powders to increase their stability and shelf-life. The dissolution process of some lyophilized some proteins is complex. In addition, proteins generally considered to be readily soluble may still require a number of hours to reach conformational equilibrium.

Finally, CE faces some limitations and therefore is not proposed as a stand-alone technique for the structural characterization of proteins. One main limitation is that the information provided is insufficient for the complete characterization of structures. To overcome this, orthogonal separations are implemented. Asymmetric field flow fractionation (AF4) and native polyacrylamide gel electrophoresis (PAGE) were able to provide information on the molar masses and conformations of the sub-populations present in protein samples. Both of these techniques have the potential to be coupled to CE for a more comprehensive characterization of proteins. This work opens the doors to advancements in the biomedical, pharmaceutical as well as food and nutrition industries.

CHAPTER 1: Background

1.1. An introduction to protein structure

Proteins are responsible for a wide variety of biological processes and thus vary in functionality and structural complexity.[1, 2] At the most basic macromolecular level, all proteins possess a distinctive primary structure, that is, the linear amino acid sequence.[1] This primary sequence is then able to fold into particularly stable arrangements of recurring structural patterns, referred to as the secondary structure which commonly includes configurations such as α-helices and β-sheets.[1] Higher-order proteins go on to form an overall three-dimensional arrangement of all the atoms in the protein, which is known as the tertiary structure.[1] Some proteins can then also exist in the form of multiple subunits and the overall arrangements of these subunits constitutes its quaternary structure.[1]

There is a strong and intricate relationship between the structure of proteins and the function/s they are intended to perform, as well as the efficiency at which they are able to do so. The folding mechanisms and molecular flexibility of proteins dictate their overall structure and thus are important topics to focus on in trying to understand the structure-function relationship and what exactly causes some proteins to misfold and lose their normal functioning capabilities.

Gaining a holistic understanding of the mechanisms by which proteins function in their native crowded cellular environments holds important implications in a number of different industries and research fields. In the biomedical industry, this knowledge directly impacts the division of more effective therapeutic approaches. As much as 98% of therapeutic targets for treating diseases including certain cancers, psychiatric disorders and inflammation, are proteins.[3] A greater understanding of these targets has encouraged the development of unconventional therapies with improved selectivity. On the medical level, many pathological abnormalities and diseases, such as Alzheimer's disease and cataracts, have been directly associated with protein misfolding and aggregation.[4] If we are able to fully comprehend the key

processes leading to aggregation of proteins that result in disease the opportunity to improve detection and diagnostic regimes for these diseases arises. This may well be the key to not only better combatting the onset and/or progression of these diseases but may also be a step toward finding a cure. In addition, significant focus has been placed on improving the design and production of protein-based drug carriers as they are biocompatible and biodegradable.[5] They also possess good solubility properties over a large pH range along with many other desirable features.[5] The food and nutrition industries may also benefit in two principle ways: through the improved understanding of the enzymatic processes involved in digestion, and through the evaluation of nutrition and food quality as proteins are an important component of a wide variety of foods.[6]

1.2. Structural characterization of proteins

The characterization of proteins involves the analytical study of the attributes pertaining to their physical or chemical properties.

Protein characterization is complex due to the intricacy of the molecules themselves. Nevertheless, significant advancements in techniques used in structural biology research have led to the determination of the higher order structures (refer to section 1.1) of a wide range of globular proteins.[4] These advancements have in turn led to an increased recognition of the strong structure-function relationship of proteins. It has also emphasized the need to re-address the textbook definition of protein structure, that is, that one specific amino acid sequence leads to a single functional conformation.[2] It has since been proposed that the native state of a protein is instead represented by an ensemble of interconvertible conformers in equilibrium, rather than a single structure.[2, 7] This is true especially for intrinsically disordered proteins which do not adopt stable overall three-dimensional structures, as well as proteins containing intrinsically disordered regions.[8] More than 50% of all known eukaryotic

proteins possess intrinsically disordered regions which make their structures far more difficult to solve.[2, 9] Protein structure-function relations can thus be refined to their function being more dependent on protein dynamism, that is the ability to transition between conformational states, than with any single structure.[7] In addition, many proteins also form functional oligomeric species or oligomeric intermediates during aggregate formation, further increasing the proteins structural complexity.[10] The 'overall structure' of many proteins is thus comprised of a number of variant sub-populations which influence their structural heterogeneity.

The sub-populations constituting a single protein may vary in three principal ways: size and/or molar mass (i.e., the native state is populated by numerous oligomers) or molecular configuration (i.e., subtle differences between conformers which occupy the same hydrodynamic volume), all of which manifest as distributions of these attributes. These three structural aspects are inherently linked, thus understanding the level and degree of heterogeneity (expressed in terms of dispersity) of each may lead to a more complete appreciation of the molecular flexibility and functioning of significant proteins.

1.3. Traditional protein characterization techniques

1.3.1. Non-separative techniques

Most of the groundwork on protein structure and function has been completed using techniques capable of elucidating their three-dimensional structures at the atomic level. X-ray crystallography is commonly used to determine the static three-dimensional structures of proteins. The major limitations of this technique include high-cost and difficulty in crystallizing some proteins, which makes it very much dependent on the expertise of the operator. In addition to this, generally only the most stable conformation (with the lowest energy) is represented in the precipitated

crystals used for analysis, leading to an inaccurate representation of the entire protein sample.[11]

Nuclear magnetic resonance (NMR) spectroscopy and circular dichroism (CD) are techniques capable of providing insight into multiple structures present within a protein sample in the form of an average structure. NMR spectroscopy is typically used in protein analysis for resolving the time-averaged structural ensemble of proteins, while CD is generally used as a fingerprinting tool for tertiary structure determination through assigning particular features of the spectrum to particular structural components.[12, 13] One major advantage of both these techniques over X-ray crystallography is their ability to elucidate the structure of proteins in solution and under conditions which better mimic physiological conditions. Both techniques do however face limitations such as restrictions in the upper size limit in NMR and issues related to data distortion in CD.[12, 14]

1.3.2. Separative techniques

Due to many proteins being heterogeneous in nature, the aforementioned techniques all face the issue of not identifying all of the possible conformations which may exist within a protein sample. Separation techniques offer the ability to analyze the entire sample through obtaining a distribution of specific molecular features within the protein. From this, the heterogeneity of the protein can be quantified in the form of a dispersity value of the relevant distribution.[15]

Polyacrylamide gel electrophoresis (PAGE) is one of the principal separation techniques used in the analysis of proteins, mainly applied for protein identification, evaluation of sample purity and molar mass estimation.[16-18] Separation takes place in a gel, according to the hydrodynamic volume of the analyte components within the sample as well as the percentage of acrylamide in the gel. The principal advantages of gel electrophoresis as an analytical tool are experimental simplicity and the ability to achieve high resolution separations of a number of proteins simultaneously.[19, 20]

There are many forms of PAGE, namely sodium dodecyl sulphate (SDS)-PAGE, native-PAGE and two-dimensional (2D)-PAGE, each with their own set of advantages and limitations.[19, 21] Polyacrylamide gels, in general, show strong affinity for proteins, making extraction of the protein for further analysis tedious.[16] The subsequent staining required for visualizing the bands also leads to slower analysis and causes the throughput of the technique to be limited.[17] Some stains are also known to interfere with the structure of the proteins being analyzed.

Currently, size-exclusion chromatography (SEC) coupled to multi-angled light scattering (MALS) is the most common analytical technique used to separate proteins based on their hydrodynamic volume, providing insight into the dispersity of molar mass and size distributions.[22, 23] SEC is commonly used, especially in the development of pharmaceutical proteins, for the characterization of aggregate size and content due to the high repeatability of the technique and instrumental availability.[24] It does however face a number of significant drawbacks such as those related specifically to the SEC column itself. Firstly, it can act as a physical filter, removing larger aggregates from the analyte, secondly, proteins can adsorb to the column, altering the retention volume, thirdly, proteins can deform under the pressure applied and lastly, samples can become diluted, which, along with the frits of the column, can cause the aggregates/oligomers to dissociate.[23, 25] These limitations lead to a loss of the true representation of the entire sample. They also explain the poor reproducibility of SEC with regards to molar mass determination, especially for lower molecular weight molecules.[25-27]

1.4. Approaching a novel technique to further understand heterogeneity

Extensive characterization of proteins using these techniques, along with many others, has led to an increased appreciation of the role that heterogeneity plays in the

overall structure. Given that proteins can exhibit structural heterogeneity, the need to separate different conformers is becoming of increasing interest to better understand the link between protein structure/s and function. This places emphasis on the search for techniques capable of characterizing the heterogeneity of proteins with regards to attributes not conventionally studied through the techniques outlined in section 1.3. Typically, a combination of methods based on different measurement principles are often employed complementary to one another for a more holistic characterisation.[25] To the best of our knowledge, however, there is currently no method capable of quantitatively obtaining the conformational dispersity of proteins. Free-solution capillary electrophoresis holds promise for this due to several advantages.

1.5. Free-solution capillary electrophoresis

Free-solution capillary electrophoresis (CE) is a robust polymer characterization technique used to separate polyelectrolytes based on their migration through a liquid medium under an applied electric field.[28] In comparison to traditional protein characterization techniques, CE analysis can be undertaken more rapidly, is more cost-effective and requires significantly less sample quantities with minimal preparation.[28, 29] The absence of a stationary phase also reduces one of the most common issues encountered in other separation techniques, that is, sample adsorption onto the stationary phase.[30]

1.5.1. Instrumentation and mechanism of separation

The separation in CE typically takes place in a narrow fused-silica capillary filled with a background electrolyte (BGE), usually a buffer solution. If the buffer has a high pH (>2 in water), the silanol groups on the inner surface of the capillary are negatively charged.[28] The ions in the buffer interact with the wall, forming an electrical bilayer which is composed of a fixed and a diffuse layer (Figure 1-1). The

diffuse layer contains a higher concentration of cations than the bulk solution, resulting in a 'dragging' motion of the entire sample due to attractive forces between these cations and the cathode. This motion is referred to as the electroosmotic flow (EOF) and, as neutral molecules will migrate at the same rate as the EOF, they are commonly used as EOF markers.[29] Positive molecules will be detected before the EOF marker due to their stronger attraction towards the cathode. For analytes with a net negative charge, attraction to the anode causes them to migrate slower than the EOF (thus they are detected after the marker). As the overall effect of the electroosmotic flow is stronger than the independent migration of the molecules, the net flow of the sample is still in the direction of the cathode.[31]

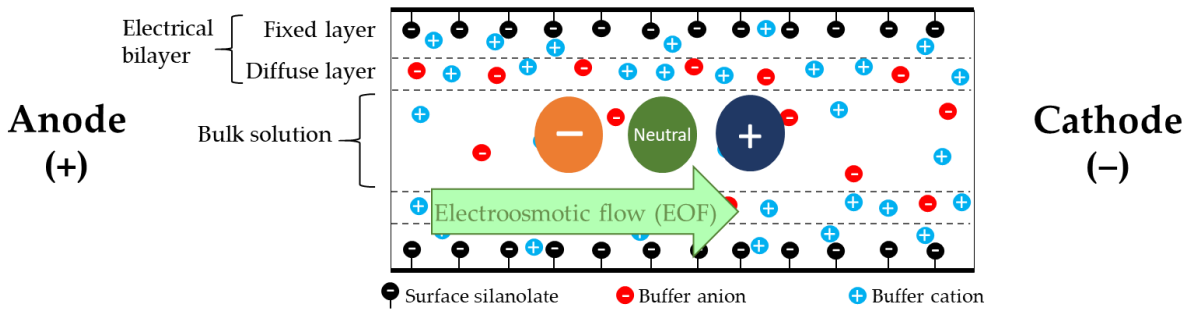

Figure 1-1 Mechanism of separation by counter EOF CE. The dense positive charge in the electrical bilayer 'drags' the entire solution in the direction of the cathode (defined as the EOF). Positive molecules migrate faster than the EOF, neutral molecules migrate with the EOF and negative molecules migrate slower than the EOF.

The mobility of a molecule through a medium under an applied electric field is dependent on the molecule's charge to friction ratio and referred to as its electrophoretic mobility.[32] Separation in CE thus occurs due to differences in the electrophoretic mobilities of the molecules within a sample.[33]

1.5.2. CE of proteins

CE embraces a large number of separation and detection modes and thus has successfully been applied to the characterization of various polymers, both synthetic

and natural.[6, 28, 33] CE has led to significant contributions to the analysis of proteins and has consequently been applied in the biomedical, pharmaceutical and food/nutrition industries. In this section, the applications of CE are briefly outlined in an attempt to emphasize the novelty of the knowledge intended to be brought about by this project. There are many more applications of CE either directly or indirectly related to protein analysis; however, this section has been limited to some of the newest and what is believed to be the most exciting developments over the past decade.

In proteomics, CE is used as a tool for diagnosing diseases through the separation and detection of biomarkers present in biological samples.[34] Examples of such cases include the diagnosis of chronic kidney disease and coronary artery disease through urinary proteomic analysis, as well as the diagnosis of phenylketonuria in newborns.[34, 35] The combination of CE and nanoparticles has enabled the detection of C-reactive protein in plasma after acute phase inflammation[35], as well as the detection of albumin in urine, which is indicative of kidney damage.[34] Protein fingerprinting of a number of bacterial species (namely *mycobacterium marinum* and two *staphylococcus* species) has also been accomplished using CE.[34, 35] In addition, high resolution CE has been used to resolve isoforms of tau protein[34] and heparan-*N*-sulfatase[35], as well as to identify proteoforms with post-translational modifications.[35] Certain protein-protein interaction inhibitors have also been screened for using CE.[34, 35]

CE has also proven useful in analyzing proteins in different foods.[6] This analysis involves determining genetic variants and detecting adulterations.[6] Specifically, it has been used for the evaluation of meat quality and processes such as heat treatment of dairy products, as well as the investigation of proteolysis in matured and fermented products (such as cheese and wine).[6]

All of the above-mentioned applications of CE involve the detection of biomarkers and the identification and quantification of certain proteins through qualitative and quantitative peak analysis. In this work, the aim is to gain further knowledge from CE separations. Analyzing the range of distributions of electrophoretic mobilities (in terms of mobility values and of dispersity) is proposed to offer more insight into protein structure and heterogeneity in native conditions.

1.6. This Project: Establishing CE as a technique to offer complementary information on protein structure

The overall purpose of this project is to apply a novel CE method to a number of already well-characterized ('model') proteins, establishing the experimental conditions for maximum separation efficiency and resolution, and finally corroborating the results with findings from complementary techniques to assist in the interpretation of results.

1.6.1. The ability of CE to separate proteins by conformational differences

The mobility of oligoelectrolytes that exhibit a low degree of polymerization (≤10)[32] has proven to be proportional to the ratio of its nominal charge and hydrodynamic volume.[32] For molecules above a certain degree of polymerization, the electrostatic friction overpowers the hydrodynamic friction, resulting in the electrophoretic mobility of these analytes having a very weak dependence on molar mass.[28]

The conditions under which molar mass independent separation occurs has been sought out in liquid chromatography in the critical conditions (LC-CC).[36] Free-solution CE of evenly charged polyelectrolytes leads to analogous separations to LC-CC, and thus the technique of free-solution CE is named here as CE in the critical conditions (CE-CC).[37]

CE-CC enables the separation and analysis of molecules based on a molecular attribute other than molar mass. It further allows a distribution of electrophoretic

mobilities inherent to that attribute to be obtained.[37] Analogous to the calculation of the dispersity of molar mass distribution commonly used in the treatment of SEC data, a calculation of dispersity of the distribution of electrophoretic mobilities was derived and successfully applied to a number of polysaccharides and synthetic polymers for which they were able to quantify the heterogeneity of composition or branching.[37] In the case of proteins, CE-CC is believed to separate purified samples based on structural differences between the individual protein molecules. A dispersity value closer to 1 indicates low sample heterogeneity, i.e., for proteins that would mean less structural variety in the conformational ensemble comprising the overall structure.[37]

1.6.2. Interpreting results from CE of proteins

To assist with the interpretation of CE data, corroboration of the results with those from other, widely-used techniques is necessary. In this work, we compare our results with those from SEC, and native-PAGE, both well-established techniques for the study of proteins. In addition, exploring another widely-used technique, however not so well-established for proteins specifically, may assist further and provide potential not yet tapped into.

1.6.3. Field flow fractionation (FFF)

Field flow fractionation (FFF) can be used to separate and characterize samples containing dispersed analytes over a broad size range.[38] The separation occurs as a result of the mobile phase (containing the analyte) flowing in a laminar fashion through a channel to which an additional field is applied orthogonally.[38] Different variants of the technique have been developed based on the nature of the applied field, however, overall FFF is considered non-denaturing due to the absence of a stationary phase. Williams, 2012 #310} Flow field-flow fractionation is a well-established technique that has previously been applied in the pharmaceutical field for the analysis of protein aggregates.[38] The microcolumn version of F4 known as

hollow-fiber flow FFF (HF5) also holds promise for the technique's applications in emerging bioanalytical fields such as protein analysis and proteomics.[38] As mentioned previously, coupling of techniques is often necessary to characterize proteins due to their complex nature. As such, the aim of coupling our method of CE-CC to FFF will enable more accurate interpretation of results which will assist in establishing CE as a routine technique for quantifying the dispersity of electrophoretic mobility distributions of proteins.

1.6.4. Proteins under study

The following proteins were selected for this study based on their isoelectric points (pI) being below physiological pH (Table A-1). This causes them to be negatively charged under the required experimental conditions (in buffer, pH ~7.4), allowing for minimal interaction with the capillary. In addition, there is a substantial amount of literature on each of their structures and functions as they have been extensively studied. They are all globular, metalloproteins.

1.6.4.1. Alcohol dehydrogenase (ADH) from *Saccharomyces cerevisiae*

ADH functions as a catalyst in the reversible oxidation of alcohols into aldehydes and is therefore well known for its role in glucose fermentation.[39] It has a homo-tetrameric structure, composed of a pair of dimers, each divided into two subunits.[39, 40] Each subunit contains a coenzyme (NAD^+) binding domain, a catalytic zinc domain as well as an additional structural zinc.[39] It is a large protein, with an approximate molar mass of 150 000 $g \cdot mol^{-1}$ and a pI of 5.4-5.7, making it well suited to analysis by CE.[39, 41]

1.6.4.2. α-Lactalbumin (α-LA) from bovine milk

α-LA is classified as a low molar mass (approx. 14 200 $g \cdot mol^{-1}$) whey protein capable of metal binding.[42-44] Its main function is to act as the regulatory subunit of an enzyme complex known as lactose synthase.[44, 45] α-LA undergoes reversible binding

to the catalytic component of the complex and subsequently promotes glucose binding to the complex, facilitating the biosynthesis of lactose.[44] Each monomeric unit is comprised of two lobes, a larger α-helical lobe and a smaller lobe composed of a small β-sheet, a small 3_{10} helix and some irregular structure.[45] These lobes are referred to as "subdomains" as they lack the typical properties of "true" domains, such as distinct functional properties and origins.[45] α-LA also contains two Ca^{2+} binding sites which can bind various metal ions in solution.[45] It exhibits a variety of conformational states and is therefore said to have enhanced molecular flexibility.[44] It also has four disulphide bonds.[46] The pI for α-LA has been determined as 4-5.[42]

1.6.4.3. Bovine serum albumin (BSA)

BSA is the major circulatory protein involved in controlling the ionized levels of calcium (and magnesium) in the blood of mammals.[47] It also serves as a transporter for a wide range of metabolites, drugs and nutrients and contributes to the colloid osmotic pressure.[47] BSA is also a reservoir of nitric oxide (signaling agent). This protein predominantly exists as a monomer with an approximate molar mass of 66 000 $g \cdot mol^{-1}$; however, dimers and trimers (and larger oligomers) do also exist under native conditions.[48-50] Overall, the protein has three Ca^{2+} binding sites, seventeen intramolecular disulphide bonds, one free sulfhydryl group and two tryptophan residues.[47, 49, 50] The pI for BSA has been published as 4.7, 4.9.[51]

1.6.4.4. Insulin from bovine pancreas

Insulin is the primary protein hormone responsible for regulating blood glucose levels.[52, 53] It is produced and stored in the pancreas as a stable, zinc-bound hexamer. Upon a change in pH, it is released into the blood serum where it dissociates into the physiologically active monomer.[54] This monomer has been published as having a molar mass of 5 700 $g \cdot mol^{-1}$ and contains one intra- and two intermolecular disulphide bonds. The structure of insulin is highly dynamic, with at least two

known stable conformations (R-state and T-state conformations).[53] The pI of native insulin is 5.3.[55]

1.6.4.5. Catalase from bovine liver

Catalase is responsible for catalyzing the conversion of hydrogen peroxide to oxygen and water, thus serving to protect cells from its toxicity.[56, 57] It predominantly exists as the enzymatically active tetramer, which consists of four identical subunits each with a molar mass of 57 000 g·mol^{-1}.[56, 58] Each subunit contains a heme cofactor bound to an iron atom.[57] The pI of catalase is 5.4.[59]

1.6.4.6. Deoxyribonuclease I (DNAse I) from bovine pancreas

DNAse I is an endonuclease, responsible for cleaving double-stranded DNA. Its activity is dependent on the presence of divalent cations, especially Ca^{2+}, Mg^{2+} and Mn^{2+}.[60] The structure of DNAse is compact and contains a carbohydrate side chain. The calculated molar mass has been published as 30 000 g·mol^{-1}.[61] Extracts of DNAse I has been shown to contain four forms of the protein (A, B, C and D), differing in carbohydrate content. The pI values are 5.2, 4.9, 5.1 and 4.8, respectively.[62]

1.7. Aims and objectives

The research question targeted by this project is whether CE-CC can be used as a method for the quantitative study of the conformational heterogeneity of proteins. The objectives are therefore as follows:

- to establish the method of CE-CC for the quantification of conformational dispersity of globular proteins,
- to explore the potential of CE to monitor the dissolution of lyophilized proteins and assess the conformational equilibrium of the dissolved molecules,

- to compare the CE data to that obtained from complementary techniques to facilitate the interpretation of the CE result,
- to briefly explore the potential of a 2D separation involving a well-established method of gel electrophoresis and CE-CC.

CHAPTER 2: Characterization of Protein Structures from the Distribution of Electrophoretic Mobilities: A Comparative Study with the Distributions of Molar Masses

2.1. Introduction

Protein structural characterization provides an insight into the physical and chemical properties that underlie the functions and properties of these essential biological molecules. The advent and subsequent improvement of characterization techniques with molecular and/or atomic resolution have advanced the understanding of many aspects of protein structure and function. These techniques include but are not limited to mass spectrometry (MS),[63, 64] solid- and solution-state nuclear magnetic resonance (NMR) spectroscopy,[65, 66] X-ray crystallography,[67] electron microscopy,[68] analytical ultracentrifugation (AUC)[69] and size-exclusion chromatography (SEC) coupled with various light scattering techniques[70]. For monomeric proteins, i.e., those which exhibit minimal heterogeneity in terms of size/molar mass, these methods can generally provide well resolved and sensitive measurements. However, there are many important proteins which do not assume a structure with a narrow unimodal distribution.[71, 72] In such cases, the pragmatic approach is to select multiple high and low atomic resolution characterization methods that provide complementary or orthogonal structural information. Often, the use of multiple characterization methods to understand the different aspects of protein structure is applied to further elucidate the interplay between a protein's tiered structure and, ultimately, its function.

When considering the many facets of a protein's structure, the occurrence of heterogeneity is an increasingly interesting physical property. Unfortunately, however, many of the aforementioned biophysical techniques do not allow for the accurate and complete characterization of heterogeneity.[73] An important and often underappreciated notion is that a protein can exhibit multiple forms of heterogeneity, namely with regard to molar mass, size, structure and composition (the latter largely related to post-translationally modified proteins), which can manifest in uniquely disperse distributions.[73-76] While molar mass, size and structural

features are indeed inherently linked, i.e., as a protein oligomerizes, the level of dispersion of these discrete distributions can potentially be vastly different for either physical and/or functional reasons. Currently, the most common method for investigating heterogeneity within a protein sample is SEC coupled to a multi-angle light scattering (SEC-MALS) detector. This allows for a distribution of molar masses to be obtained and the heterogeneity thereof can be quantified in terms of dispersion. The standard measure of dispersion for molar mass distributions is the dispersity ($Đ$), previously named polydispersity index (*PDI*), as given by the weight-average molar mass (M_w) over the number-average molar mass (M_n).[77] Alternatively, the standard deviation (σ) can be used to express the degree of heterogeneity, giving an intuitive idea of the variation that is present in the sample.[78] The general expression for the dispersity of weight distributions is given by $D(W(A),b,c)$, which is defined as the dispersity of the weight distribution of the variable A as a function of the order of moments (b) with respect to a reference (c).[37]

Another often underutilized dispersity in protein investigations can discern distributions in size, i.e., compact or extended isoforms, expressed through differences in a protein's hydrodynamic radius ($D(W(R_h))$. Dynamic light scattering (DLS), either in batch mode or coupled to SEC (SEC-DLS), is most commonly used for the determination of the weight- and number-average hydrodynamic radius (R_{hw} and R_{hn}, respectively) via the autocorrelation function.[79, 80] These existing methods probe the dispersity of molar mass and size distributions, however, until recently, no method has been available to provide a measure of the dispersity of structural and/or compositional distributions.

One characterization technique that has been used extensively in proteomics for separation and biomarker detection purposes, but rather neglected of-late with regard to protein characterization, is free-solution capillary electrophoresis (CE).[81-83] Separation in CE occurs as a result of the charge-to-friction ratio of the analyte

molecules.[84] CE has a specific intrinsic feature that has long been underappreciated, however happens to be exceptionally attractive for the characterization of a wide range of polyelectrolytes (including proteins). CE in the critical conditions (CE-CC) leads to separations analogous to liquid chromatography in the critical conditions (LC-CC), and thus differs from traditional hydrodynamic based separation methods. The separation in CE-CC occurs on the basis of structural and/or compositional feature(s) of a polyelectrolyte and is independent of molar mass.[28] This molar mass independent separation has been repeatedly demonstrated in the literature for various polyelectrolytes including denatured protein-SDS complexes[85], DNA[86], poly(styrene sulfonate)[84] and polyacrylate[87, 88]. While a CE experiment yields separation data in the form of detector signal against the migration time, this can be transformed into a distribution of electrophoretic mobilities.[89] Contrary to migration time, the electrophoretic mobility (μ) is independent of variations in the electroosmotic flow as well as experimental parameters such as capillary length and voltage. Contrary to UV signal, the distribution is independent of the different relative velocities of the analytes. CE-CC allows for quick and accurate determination of the weight-distribution of electrophoretic mobilities, $W(\mu)$. These distributions are reflective of a distribution in a particular structural and/or compositional feature, such as the acetylation of chitosan[90, 91] and the branching of poly(sodium acrylate)[92]. Expressions of the average mobilities and dispersities of the $W(\mu)$ have been recently derived.[37] This novel approach of characterization is yet to be applied to proteins. When using CE-CC to characterize a protein devoid of post-translational modifications, any variance in the distribution of the electrophoretic mobilities should be reflective of differences in structure or otherwise referred to as conformation. This effectively allows the study of structural, or conformational, differences between related or unrelated proteins, especially their heterogeneity, provided they satisfy the conditions of CE-CC.

In this work we compare the respective electrophoretic mobility distributions, average electrophoretic mobilities and dispersities of three proteins that have been subjected to extensive structural characterization. We also compare these results to their molar mass distributions acquired from SEC-MALS. This information, along with the extensive literature surrounding important conformational features of these model proteins, allows for a holistic and robust study into molar mass and structural distributions inherent to a protein's quaternary structure.

2.2. Experimental section

2.2.1. Materials

All lyophilized protein samples including α-lactalbumin from bovine milk (α-LA), albumin from bovine serum (BSA) and alcohol dehydrogenase from *Saccharomyces cerevisiae* (ADH), were purchased from Sigma-Aldrich and used without further purification. All other reagents, unless otherwise stated, were also purchased from Sigma-Aldrich. All solutions were prepared using Milli-Q water (18 MΩ cm^{-1}, 0.22 μm membrane filter). Experiments were conducted in aqueous sodium phosphate buffers. A chloride-containing phosphate buffer (NaPi) was produced as a 200 mM stock by mixing NaH_2PO_4 with the conjugate base to a pH ~7.8. This was adjusted down to pH 7.3 or 7.4 using HCl. A 200 mM stock solution of chloride-free phosphate buffer (PB200) was also prepared by mixing disodium hydrogen phosphate (Na_2HPO_4, 1.42% (w/v)) with sodium dihydrogen phosphate (NaH_2PO_4, 1.56% (w/v)) and titrating with NaOH to pH ~7.4. All buffers were sonicated for 5 min prior to their use. The CE background electrolyte (BGE) was 5 mM NaPi or PB (pH ~7.4). The eluent used for SEC was 50 mM NaPi at pH 7.4 with 100 mM NaCl. The concentrations of protein samples were estimated using the protein's extinction coefficient ($E^{1\%}$) value (Table A-1) and measured using an ultra-low volume spectrophotometer (NanoDrop (ThermoFisher Scientific Inc.)).

2.2.2. Determination of d*n*/d*c* values for SEC

For each protein sample, a concentration series consisting of 10 points approximately between 0.5 $g \cdot L^{-1}$ and 5 $g \cdot L^{-1}$ were prepared from a stock in the filtered (0.22 μm, Chromfilter) SEC-MALS eluent (50 mM NaPi + 100 mM NaCl, pH 7.4) which had been sonicated for 5 min. The refractive index (RI) of the protein solutions at each concentration was determined using an ultra-low volume refractometer (J357, Rudolph Research Analytical) at 25 °C. The data was plotted as RI vs. concentration ($g \cdot mL^{-1}$) and the d*n*/d*c* value was taken as the slope of the linear fit.

2.2.3. Capillary Electrophoresis

CE and pressure mobilization (PM) experiments were conducted on an Agilent CE 7100 (Agilent Technologies) and a Sciex P/ACE MDQ CE system (AB Sciex Separations, Mount Waverly, Australia).[93] A fused silica capillary with an internal diameter of 50 μm was used for all experiments. For experiments conducted on the Agilent system, the total length was 40 cm (with an effective length of 31.5 cm) and for those conducted on the Sciex system, the total length was 40.2 cm (with an effective length of 30.2 cm). Capillary pre-treatment consisted of a 10-min flush with 1 M NaOH, 5-min flush with 0.1 M NaOH, 5-min flush with Milli-Q water and a 5-min flush with the background electrolyte. This pre-treatment was also used before each new protein was analyzed. Between individual experiments, the capillary was flushed for 5 min with the background electrolyte. The capillary post-treatment was a flush for 5 min with 1 M NaOH, 10 min with Milli-Q water and 10 min with air for capillary storage. All protein samples were prepared to a concentration of ~ 1 $g \cdot L^{-1}$. The specific concentration for each is shown in Table A-2. Protein samples were injected hydrodynamically into the capillary with the injection pressure set to 30 mbar for a duration of 10 s on the Agilent CE, and 34 mbar for a duration of 8.7 seconds on the Sciex instrument. The approximate injection volumes were estimated using the Poiseuille equation for each instrument as 13.42 nL and 11.44 nL,

respectively.[94] CE separation was conducted at 30 kV while PM experiments were completed using 50 mbar of pressure. Oligo(sodium acrylate) in 25 mM sodium borate buffer (pH 9.3) was used to validate the capillary.[87, 95] The neutral molecule dimethyl sulfoxide (DMSO) was used as a mobility marker in CE experiments at 0.4 % (v/v) but was absent for PM experiments. All data was recorded at 191 nm (bandwidth of 2 nm) and treated using Origin (versions 8.5 and 9) to calculate the distribution of electrophoretic mobilities.[37, 89]

2.2.4. Size-exclusion chromatography with light scattering detection

SEC–MALS experiments were performed using a DAWN HELEOS 8 (Wyatt Technology Corp., Santa Barbara, CA, USA) laser light scattering device with a Quasielastic Light Scattering (QELS/DLS) module (Wyatt) housed in the same unit and an Optilab rEX (Wyatt) refractive index detector connected in series. Both detectors were thermostatted at 25 °C and the eluent was degassed online with a Waters In-Line Degasser AF. Protein samples were made to a concentration of 2 $g \cdot L^{-1}$. Sample separation was achieved using a GE Superdex 200 10/300 GL SEC column equilibrated with the sample buffer using a 100 µL injection loop. Normalization of the MALS detectors was achieved using the monomeric fraction of BSA at 10 $g \cdot L^{-1}$ in the aforementioned equilibration buffer. Data acquisition and processing were completed using ASTRA (5.3.4) software (Wyatt). Further data treatment and plotting was completed using Origin 7.

2.3. Results and discussion

Adsorption is a plague in separation and characterization of macromolecules, both in capillary electrophoresis[96] and size-exclusion chromatography[97-99]. One way to monitor adsorption in CE, however, is to do complementary pressure mobilization experiments. Pressure mobilization involves applying pressure to induce the migration of an analyte through the capillary. The absence of an electric field means

no separation takes place, but the presence of band broadening leads to a Gaussian peak if no significant adsorption takes place on the capillary wall.[28, 100-102]

2.3.1. Pressure mobilization

In this work, pressure mobilization (PM) experiments were performed to monitor whether any of the proteins adsorbed onto the inner capillary wall. Adsorption is detected by the presence of tailing.[96] For each of the proteins studied, a single, symmetrical peak is observed with little discernable tailing (Figure 2-1). The lack of visible skewness therefore suggests that little to no adsorption onto the capillary surface has occurred.[102] For comparison purposes and to ascertain this observation statistically, the peaks were fitted with a Gaussian function. The coefficient of determination (R^2) values are indicative of the quality of the Gaussian fit. For all three proteins analyzed, the average R^2 value is above 98%, again indicating virtually no tailing and thus means very little adsorption (if any) has occurred during PM. The R^2 values for each protein obtained from the two instruments are shown in Table A-2.

PM is also used in conjunction with CE to determine the relative recovery of the sample. The percentage relative recovery for each protein across instruments is given in Table A-2. For each protein, regardless of which instrument the experiments were done on, the area from PM peaks are larger than for CE peaks. The recovery varies with each protein and no trend is seen in which instrument consistently gives a better recovery. For BSA, the recovery is high enough as the percent difference between the PM and CE peak areas fall within the 15% margin of error. The recovery for ADH and α-LA, however, falls above this error. The loss of sample for these to proteins is likely not adsorption to the capillary wall, as adsorption is more likely in PM. This means that if adsorption was the cause, the peak areas in CE would be larger than the PM peak areas. As this is not the case, the sample loss here is likely

due to sample degradation in the detection window or a result of photodegradation under the influence of the electric field.[103]

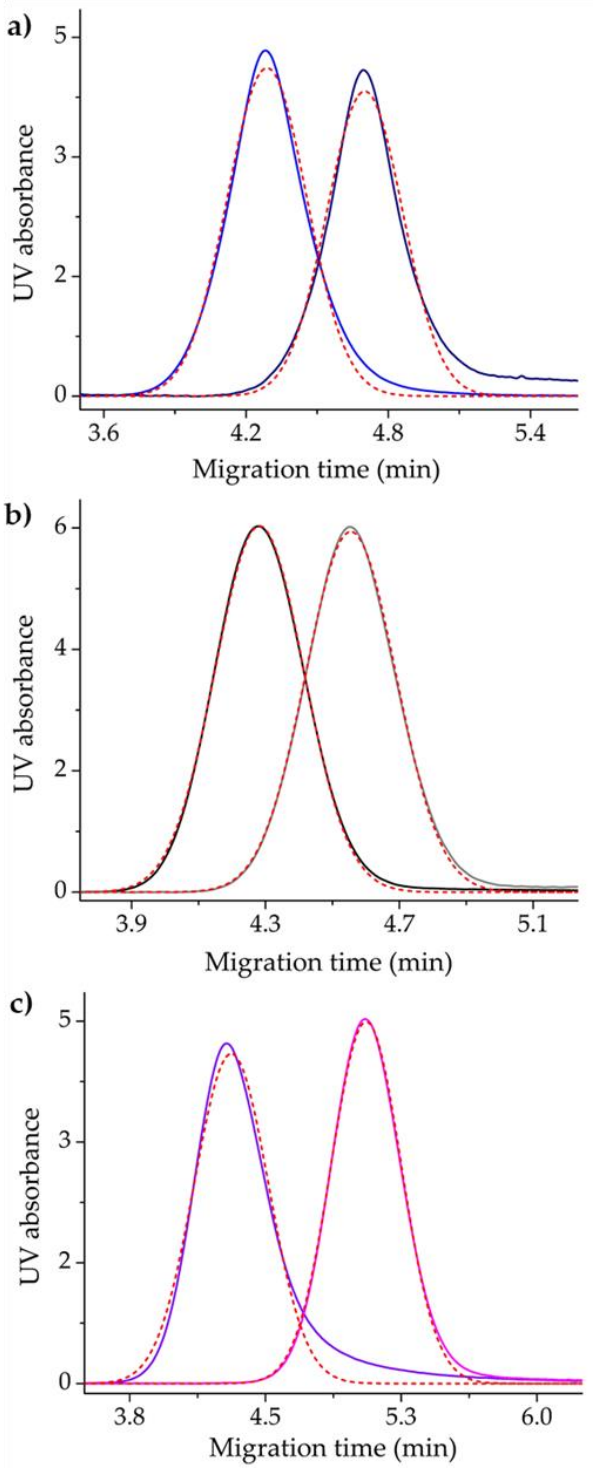

Figure 2-1. Representative pressure mobilization results from Sciex (left) and Agilent (right) instruments displaying the fits of the Gaussian function (red, dashed) for a) ADH, b) α-LA, and c) BSA.

2.3.2. Separation by capillary electrophoresis and size-exclusion chromatography

The raw data from CE and SEC experiments was transformed to produce distributions of electrophoretic mobilities and molar masses, respectively. The dispersities for each respective distribution were calculated, using M_w/M_n for SEC data, and an analogous calculation for CE data (Table 2-1).[37]

2.3.2.1. Bovine serum albumin

CE of commercial BSA revealed a monomodal distribution of electrophoretic mobilities (Figure 2-2, a and b) and a multimodal molar mass distribution (Figure 2-2, c). The broadness of the peaks and presence of at least one discernible shoulder (in both CE instruments), however, still suggests that this sample exhibits a degree of heterogeneity in terms of conformation. This is also quantitatively reflected in the value of $D(W(\mu),1,0)$ for BSA (1.006), which is considerably higher than that of linear poly(sodium acrylate) (1.000013), a homopolymer.[37]

A number of previous studies using commercial samples of BSA have indicated the existence of BSA in different oligomeric states.[104] This is evident in the SEC-MALS results presented here (Figure 2-2, c), where the multimodal distribution of molar masses has at least three distinct modes with values that correspond to monomeric, dimeric and trimeric forms of BSA.[104, 105]

Due to the marked visual difference in the distribution of molar masses and electrophoretic mobility for BSA, it is unlikely that the conformational heterogeneity apparent in the distribution of electrophoretic mobility is simply a result of oligomers present in the sample. Instead, these results suggest not only molecules of varied sizes exist in the sample, but that different conformations for each oligomer might also exist, increasing the overall conformational heterogeneity of the sample.

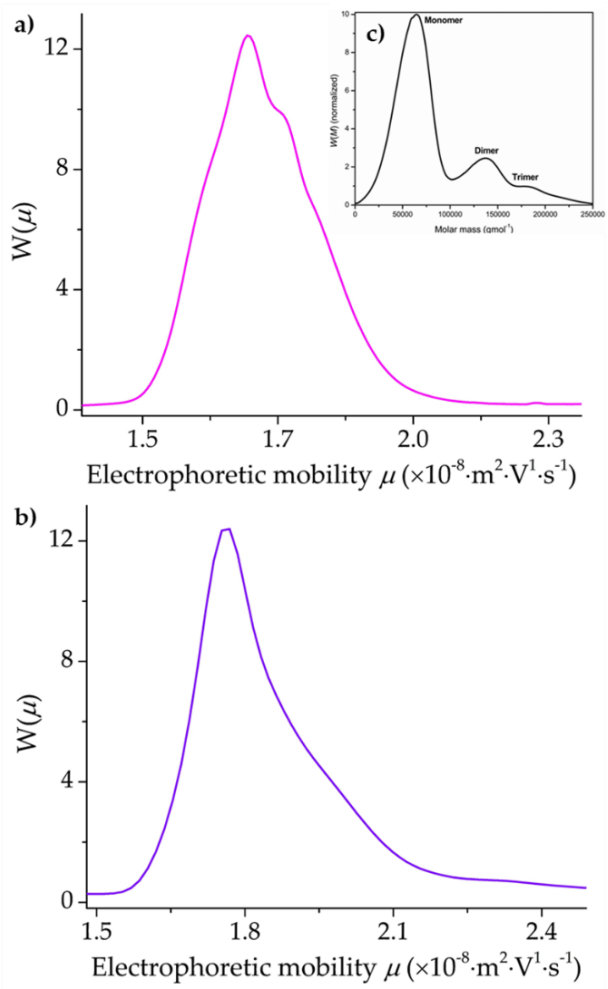

Figure 2-2. Electrophoretic mobility and molar mass distributions of BSA obtained by CE (a and b, corresponding to Agilent and Sciex instruments, respectively) and SEC-MALS (c).

2.3.2.2. α-Lactalbumin

The multimodal distribution of electrophoretic mobility for α-LA (Figure 2-3, a and b) indicates a more marked sample heterogeneity. A different separation is observed with SEC-MALS, however, where the distribution of molar masses (Figure 2-3, c) is monomodal. The molar mass value obtained is approximately 14,000 g·mol^{-1}, which corresponds to the monomeric form of the protein.[42, 43] Thus, it can be inferred that the heterogeneity detected in CE is independent of molar mass and could instead be attributed to differences in the conformation of the protein.

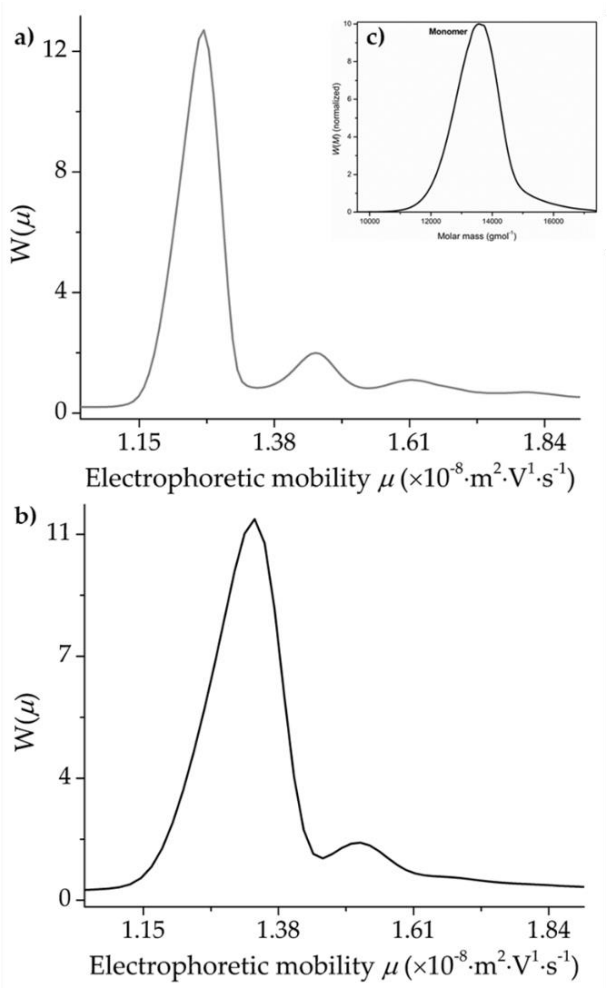

Figure 2-3. Electrophoretic mobility and molar mass distributions of α-LA obtained by CE (a and b, corresponding to Agilent and Sciex instruments, respectively) and SEC-MALS (c).

α-LA has a high affinity for binding Ca^{2+}.[45] This binding is crucial in stabilizing the holo-form of the protein against denaturation.[44, 45] The electropherograms presented here are that of a calcium-depleted sample of α-LA. The presence of multiple peaks is thus likely due to either of two reasons, or a combination of both. The peaks may indicate different conformers of the apo (Ca^{2+}-unbound) form of the protein as it has been documented that at ambient temperature and low ionic strength, apo-α-LA undergoes a transconformation to a molten-globule state, with no fixed tertiary structure.[45]

Instead or in addition to the above explanation, another may be that the peaks are representing the holo-protein, bound to metals other than the preferential Ca^{2+}. This is because the two Ca^{2+}- binding sites of α-LA are able to bind a number of different metal ions in solution.[45] Thus, the conformational heterogeneity could be due to the protein binding to trace amounts of metal ions present in the water used for the sample preparation, resulting in a number of different conformational states depending on the ligation state.[45] To determine whether this is the case, further experiments are necessary, such as the addition of EDTA to inhibit protein-metal binding.

2.3.2.3. Alcohol dehydrogenase

The distribution of electrophoretic mobilities for ADH is monomodal, however the presence of shoulders on either side of the main peak (Figure 2-4, a and b) suggests a number of conformations are present in the sample. According to literature, ADH from lower eukaryotes exists predominantly as a tetramer.[39] This is reflected in the SEC-MALS data (Figure 2-4, c), as the molar mass distribution yields a monomodal peak. The approximate molar mass is 150,000 $g \cdot mol^{-1}$, which is congruent with previous studies.[106]

The conformational heterogeneity of ADH observed in CE may arise due to a number of reasons. Firstly, there are at least four known variants of the monomeric subunit constituting the overall tetrameric unit. These are simply referred to as subunits A, B, C & D and although they are considered identical, superpositioning of their crystal structures revealed slight differences in the conformation of each.[39, 40] These subunits can form a number of different dimers and subsequently tetramers, resulting in greater conformational heterogeneity.[39] In addition to this, regions of disorder are also known to exist, especially in subunit D, which could also increase the number of different structures present in the sample.[39]

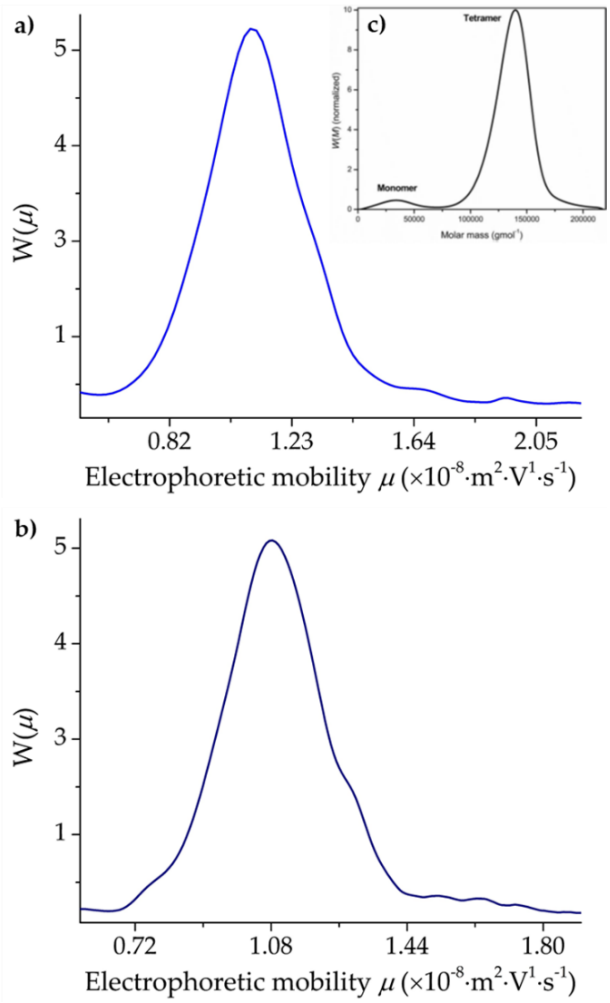

Figure 2-4. Electrophoretic mobility and molar mass distributions of ADH obtained by CE (a and b, corresponding to Sciex and Agilent instruments, respectively) and SEC-MALS (c).

2.3.3. Dispersity of the electrophoretic mobility and molar mass distributions

As mentioned previously, in addition to qualitatively comparing the distributions of electrophoretic mobilities and molar masses, the heterogeneity in terms of molar mass and conformation has also been quantified for these proteins. Heterogeneity is quantitatively expressed in terms of the dispersion of the distribution of a molecular attribute. Dispersion may be represented by a dispersity value, which is a relative measure, or in terms of standard deviation, which is an absolute measure. The

dispersity of the weight distributions of electrophoretic mobilities and molar masses are given in Table 2-1.

Table 2-1. Summary of characterization by CE and SEC. Dispersion of the samples are shown as $D(W(\mu),1,0)$, $D(W(\mu),2,0)$, and D_σ in terms of electrophoretic mobility, and as $Đ$ in terms of molar mass. The error was estimated using the standard deviation of n=9 for CE and n=3 for SEC.

	CE				SEC		
Protein	**Dispersity ($D(W(\mu),1,0)$)**	**Dispersity ($D(W(\mu),2,0)$)**	**Standard deviation (D_σ)**	**Weight average mobility (μ_w)**	**No. of peaks**	**Weight average molar mass (M_w) ($g \cdot mol^{-1}$)**	**Dispersity ($Đ$) (M_w/M_n)**
ADH	1.034 ± 0.008	1.034 ± 0.009	2.14E-09 ± 3.05E-10	1.16E-08 ± 3.93E-10	1	137 617 ± 6214	1.02 ± 0.009
α-LA	1.011 ± 0.002	1.012 ± 0.003	1.54E-09 ± 1.94E-10	1.41E-08 ± 6.50E-10	1	13 884 ± 1414	1.03 ± 0.033
BSA	1.006 ± 0.001	1.007 ± 0.002	1.48E-09 ± 2.91E-10	1.78E-08 ± 8.10E-10	1	67 549 ± 2392	1.2 ± 0.037
					2	126 700 ± 1438	
					3	190 010 ± 2434	

Dispersity values are shown as $D(W(\mu),1,0)$ and $D(W(\mu),2,0)$, which are calculated as ratios of different orders of moments and use zero as a reference.[37] Standard deviation is related to the dispersity and the weight-average of variable A (in this case either molar mass or μ) is used as the reference.[37] For any distribution, at least 75% of the values fall within 2 standard deviations of the mean.[78] Either of these values can be used as a quantitative measure of conformational heterogeneity.

Although both standard deviation and dispersity values of electrophoretic mobility distributions can be used as measures of heterogeneity, their independent relevance will depend on the samples being analyzed and the properties being assessed. For the three proteins studied here, the weight-average mobilities for all three are close (Figure 2-5), meaning that they all exhibit similar charge to friction ratios.

As the mobilities for all three proteins are similar, the same trend is observed in terms of the dispersion expressed as dispersity or standard deviation.

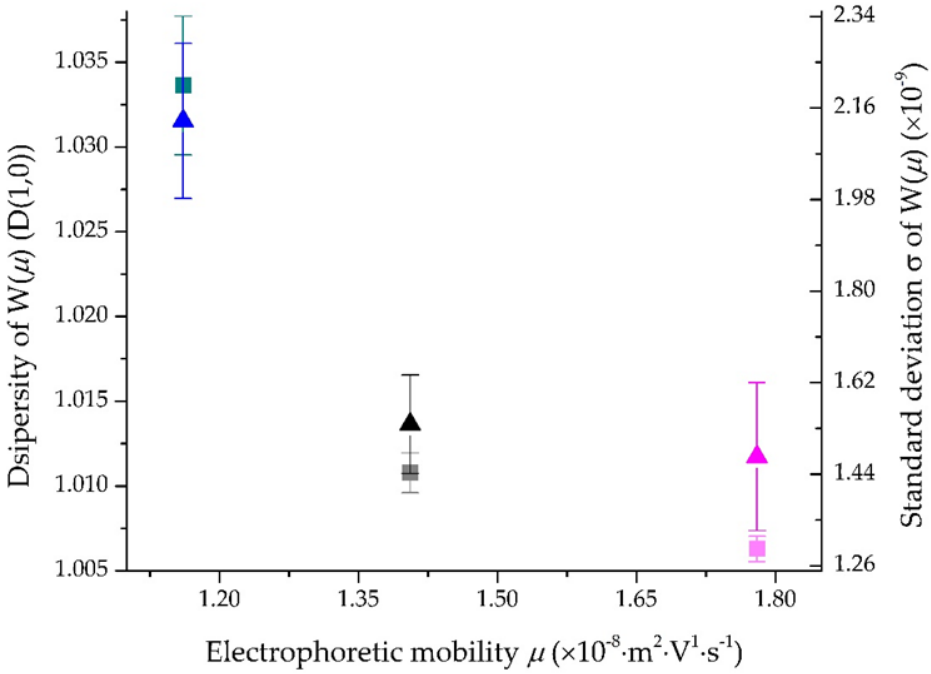

Figure 2-5. Dispersity values shown as $D(W(\mu),1,0)$ (squares) and D_σ (triangles) for ADH (turquoise and blue), α-LA (gray and black) and BSA (light pink and dark pink) against their weight-average electrophoretic mobilities.

2.3.4. Heterogeneity in terms of molar mass and conformation

Often in protein characterization studies, the heterogeneity related to molar mass is relayed in terms of a dispersity value (M_w/M_n) of the molar mass distribution. The weight-average molar mass (M_w) and $Đ$ for ADH, α-LA and BSA were determined through SEC-MALS and these values are displayed in Table 2-1. From Figure 2-6,

BSA has a significantly higher $Đ$ than ADH and α-LA. It also has the highest weight-average electrophoretic mobility. Conversely, it has the lowest $D(W(\mu),1,0)$ of all three proteins. ADH shows the opposite trend as it has the lowest $Đ$ and weight-average electrophoretic mobility, yet the highest $D(W(\mu),1,0)$.

As mentioned previously, the molar mass, size and conformation of proteins are all linked, however some proteins may exhibit different levels of heterogeneity for each of these attributes. This is evident in the results for BSA and ADH (Figure 2-6). Once again, it is important to note that although the dispersity of electrophoretic mobility distributions are obtained through a calculation analogous to that used to determine the dispersity of molar mass distributions, these two dispersity values are not related.

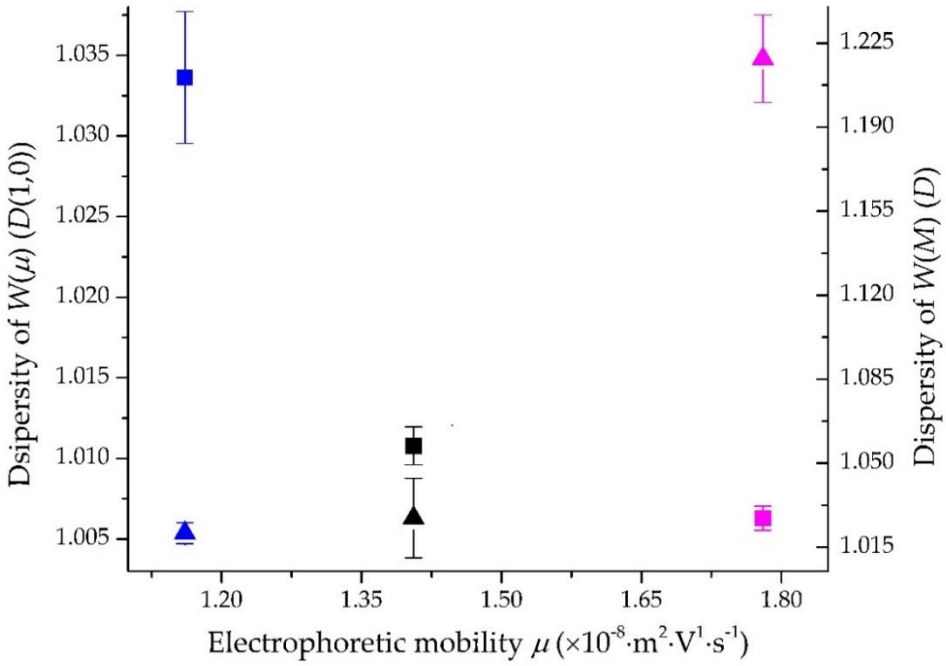

Figure 2-6. Dispersities of the distributions of: electrophoretic mobility ($D(W(\mu),1,0)$) (squares) and molar mass distributions ($Đ$) (triangles) for ADH (blue), α-LA (black) and BSA (pink) against their weight-average electrophoretic mobilities.

ADH exhibits a low dispersity in terms of molar mass relative to BSA (triangles, Figure 2-6). This agrees with literature in terms of the predominant form of this protein in solution being the tetramer, indicating that the majority of the protein would have a narrow distribution of molar masses. This agrees with the SEC-MALS data, where a large peak shows the bulk of the sample having a molar mass around 140,000 $g \cdot mol^{-1}$ (Figure 2-4, c). The dispersity of electrophoretic mobility distribution, however, is higher in comparison to the other two proteins (squares, Figure 2-6). This in fact supports the above interpretation of the electropherogram for ADH (Figure 2-4, a and b), which suggests that the tetramer is capable of coming together in multiple different ways, leading to different conformations of the protein, all exhibiting the same hydrodynamic volume.

BSA on the other hand and as stated before, is known to exist in multiple oligomeric states in solution, and the SEC results indicate the presence of at least three states differing severely in molar mass. The high dispersity of molar mass relative to ADH is thus explained, however interestingly, the dispersity of its electrophoretic mobility distribution is the opposite. BSA has a much lower dispersity in terms of conformation when compared to ADH. The separation in CE is dependent on the charge to friction ratio of the molecules being separated. This means that the different oligomers of BSA may have very similar conformations in terms of compactness of the molecules and exposed regions. The electrophoretic mobility distributions themselves (Figure 2-2, a and b) are not symmetric (distinct shouldering) and thus indicates separation is occurring despite very similar mobilities of the oligomers.

The dispersities of both molar mass and electrophoretic mobility distributions for α-LA do not vary as distinctly as is the case for the previous two proteins. The low dispersity of molar mass distribution agrees with the monomodal peak from SEC-MALS. The increased dispersity of electrophoretic mobility of α-LA relative to that

of BSA makes sense when considering the fact that this protein has enhanced molecular flexibility as discussed previously.[44] These figures all support the notion that samples can exhibit heterogeneity of several attributes not necessarily only related to molar mass in proteins. It also proves the capability of CE-CC to separate proteins based on an attribute not conventionally considered and this occurs independent of molar mass.

2.4. Conclusions and future work

The findings from this study offer insight into the potential of CE in providing additional information on the properties that influence or dictate the overall structure/s of proteins. This may prove very useful in better understanding concepts such as molecular flexibility and aggregation. These concepts are not yet fully understood, despite them being extremely important as they directly relate to the proper functioning of many proteins. As it has been extensively documented that malfunctioning of many proteins lead to diseases, the search for experimental techniques to better understand the structure-function relationship has continued throughout the decades. CE offers many advantages in this respect, in that not only does it provide insight into the heterogeneity (of proteins) with regard to features not conventionally studied, but it is also a far less time consuming, robust and cost-effective analytical technique. Appropriate treatment of CE-CC data enabled the quantification of the conformational heterogeneity of three already well-characterized proteins. A numerical representation of the mobility distributions for each protein has also been calculated and this enables comparisons between the different protein samples to be made. Dispersity values expressed as $D(W(\mu),1,0)$ and standard deviation of $W(\mu)$ were in good agreement in terms of the trend observed. Additionally, CE can be used to monitor processes such as dissolution, and chemical reactions such as fermentations.[29, 100]

Establishing CE as a protein characterization technique to be used complementarily or in coupling to other techniques, as detailed in Chapter 4, provides the opportunity to gain a more holistic understanding of protein structure and function. Furthermore, this technique can be applied to more complex proteins that have not been characterized as well as those investigated here, to gather some preliminary structural information in a timelier manner.

CHAPTER 3: Monitoring the Dissolution and Conformational Equilibrium of Proteins using CE-CC

3.1. Introduction

The molecular environment surrounding a protein directly affects the implementation of its biological functions.[107] As mentioned previously, one especially advantageous feature of CE is the ability to analyze proteins in free solution, meaning conditions can be altered to better mimic the aqueous physiological environments that most globular proteins freely exist in. This is an imperative aspect in the pursuit of gaining a more comprehensive understanding of the native structures of proteins and how it relates to their biomolecular activity. Selecting an appropriate solvent for protein analysis is crucial as the interactions between the protein residues and solvent molecules may have a significant impact on both protein solubility and overall structure.[107] In the following sections (3.1.1. and 3.1.2.), two related concepts, concerning solubility and structural equilibrium, are initially discussed separately. Thereafter, the link between them is emphasized in section 3.1.3, forming a hypothesis that will be addressed in this chapter.

3.1.1. A basic definition of protein solubility

"Protein solubility is a thermodynamic parameter defined as the concentration of protein in a saturated solution that is in equilibrium with a solid phase under given conditions."[108] The solubility of proteins in aqueous solutions may vary from hundreds of milligrams per milliliter to being almost completely insoluble.[108] Depending on the nature of the protein, solubility is variable and is determined by a number of intrinsic and extrinsic factors. At the molecular level, globular proteins, being amphipathic, self-assemble such that the non-polar groups are typically confined within the three-dimensional structure, and only polar (or charged) groups are on the surface and in contact with the surrounding aqueous environment. This intrinsic property of globular proteins is generally what renders them soluble in water and most other polar solvents.[1] Extrinsic factors influencing a proteins solubility include pH, temperature, salt concentration and other solvent additives.[1,

[108] Extensive research into protein solubility is important in the pharmaceutical industry for the improvement of pharmaceutical proteins, as well as the biomedical industry as low solubility has been implicated in various diseases.[108]

3.1.2. Conformational equilibrium of proteins

As discussed in chapter 1, many proteins are dynamic and exist as an ensemble of conformers in equilibrium.[7] Extensive efforts have thus gone into the development of techniques that enable these dynamic systems to be understood and described in terms of ensembles of structures.[109] A major challenge in doing so lies in the difficulty in quantifying and characterizing the various conformations that may co-exist in equilibrium, as well as the functionally important motions responsible for these shifts.[109] The ensembles can be interconvertible and may vary in global orientation of the proteins functional domains (i.e., in the case of the tyrosine kinase enzyme, Hck)[110], or in the shape/size of the overall molecular assemblies (as in the case of BSA)[111].[109] The processes driving the conversion, the magnitude of the changes and the rate at which they occur are all factors which may crucially impact the overall functionality of the protein. The reorientation of a proteins' domains is often involved in its regulatory functions[109, 112]. The larger scale conversions (association and dissociation of multimers) are often involved in storing and subsequently releasing proteins from a more stable multimer.[54] In other instances, however, for proteins without flexible domains, the binding of certain ligands may also be regulated by multimer formation/ dissociation.[111] Being able to attain a dynamic conformational equilibrium is therefore essential for proteins to adopt the necessary conformations in order to meet the requirements of the cell.

3.1.3. Impact of lyophilization on the stability, solubility (and activity) of proteins

All proteins used throughout this project were acquired and stored as lyophilized powders. Proteins are often lyophilized in an attempt to increase their stability for

shipping/distribution and storage purposes. This also offers the end-user an increased shelf-life of the product.[113] There is, however, evidence that many proteins do not exhibit the same conformation in their lyophilized states as they do in aqueous solutions, suggesting that structural transitions occur as protein molecules are transferred from a solid phase into a liquid phase during the process known as dissolution. [107, 114] This is supported by the fact that upon the removal of water during lyophilization, many of the non-covalent interactions (namely the hydrogen bonds and hydrophobic interactions) responsible for stabilizing the three-dimensional structures of proteins are disrupted.[1, 113] The absence of water molecules therefore often leads to intermolecular aggregation as the amino acid side groups interact in order to reach the most thermodynamically stable state.[113] In the presence of moisture however, lyophilized proteins may undergo irreversible aggregation which has been ascribed to covalent intermolecular interactions such as disulphide interchange and the formation of intermolecular β-structures.[113] The reversibility of aggregation upon rehydration thus depends on the interactions that caused the aggregation in the first place. The use of disaccharides as protectants during the lyophilization process is proposed to increase the stability of lyophilized proteins to retain conformational integrity and activity during storage.[113]

It is therefore crucial when working with lyophilized proteins to be aware of the interactions that may occur between the protein residues and solvent molecules, as aggregation may impact the solubility. Additionally, incomplete dissolution may lead to an inaccurate representation of the protein structure, stability and equilibrium when re-suspended into an aqueous environment. In addition to this, even if a protein is considered to be dissolved, a sudden change in the molecular environment might induce non-permanent interactions which may impact the equilibrium for a period of time.

The proposed hypothesis for this chapter is that as proteins go through the process of dissolution, they will undergo conformational changes until sufficient time has passed at which point the concentration remains constant, and a stable equilibrium in terms of conformation is achieved.

3.1.4. Employing the method of CE-CC to monitor the dissolution and conformational equilibrium of proteins

In Chapter 2, protein dissolution was qualitatively assessed through visual observations prior to CE analysis, where a transparent solution was assumed to be dissolved. Of the six lyophilized proteins acquired for this project, all but one resulted in transparent solutions when added to a phosphate buffer. Insulin instead formed a cloudy suspension which may be explained by its low solubility at neutral pH[115], which has stipulated on the product information sheet provided by the supplier. To compensate for this, more time was provided for it to dissolve prior to analysis. Upon standing for approximately one hour, however, sedimentation occurred, leading to a visible pellet at the bottom of the Eppendorf tube in which the sample was prepared (Figure 3-1).

Figure 3-1 Approx. 1 g (±0.001) of lyophilized insulin in 1 μL of 5 mM phosphate buffer (pH 7.4) after 1-hour incubation at ambient temperature. The presence of sediment in the form of a white pellet at the bottom of the tube indicates low solubility.

This led to the CE results for insulin being of questionable reliability and thus this data was excluded from Chapter 2. It should be noted that clear and transparent solutions do not necessarily ensure complete solubility, as has been found for starch, chitosan, poly(acrylic acid) and poly(sodium acrylate).[100, 116, 117]

A number of methods are capable of providing insight into the folding and unfolding events of proteins, which is then related back to their dissolution. Examples include the use of electrospray ionization mass spectrometry based methods for the investigation of structural transitions during the dissolution of proteins and precipitation experiments for solubility estimations.[107, 108] These methods are useful for obtaining insight into the features that impact dissolution, however are not capable of accurately quantifying the extent of protein dissolution. To do so, a method of nuclear magnetic resonance (NMR) spectroscopy previously used to quantify the dissolution of starch could potentially be employed for proteins as well.[116]

CE has previously been implemented in monitoring a number of reactions, including the hydrolysis of soy proteins as well as ethanol fermentations.[29, 118] CE has also been employed for the evaluation of chitosan dissolution in different solvents.[100] To the best of my knowledge, however, CE has not been used to study the dissolution of proteins. As a consequence, the methodology utilized for the analysis of chitosan dissolution was modified for the analysis of protein dissolution. Most of the modifications were inspired by the methodology described in Chapter 2 (Section 2.2.3).

In addition to adjusting the experimental parameters, the need for a new standard was also considered. A typical standard is comprised of components with a range of known mobilities and can thus be used for the validation of the capillary and the CE instrument.[119] The standard used widely within our research team (and therefore also in chapter 2) is an oligo(sodium acrylate) in sodium borate buffer (pH 9.3).[87]

Usually, the standard is injected before and after each experimental sequence for validation purposes. During longer experiments (>8 hours), it may become necessary to validate the instrument more often. i.e., at several intervals during the experiment. For the sake of keeping as many variables as possible constant throughout the duration of the experiment, using the same BGE for the separation of the standard and the samples may assist in minimizing errors caused by technical factors.

As we explore the potentials of CE for the characterization of proteins, the primary objective of this chapter was to use the method of CE-CC to monitor the dissolution of BSA and insulin, and to draw relative conclusions regarding their conformational equilibria through the comparison of the dispersities of their electrophoretic mobility distributions.

3.2. Experimental section

3.2.1. Materials

Lyophilized powders of BSA and insulin from bovine pancreas were purchased from Sigma-Aldrich and used without further purification. All other reagents were the same as in chapter 2. Citric acid and 2-napthalene sulfonic acid were purchased from Sigma-Aldrich, and benzoic acid was purchased from ThermoFisher Scientific.

3.2.2. Standard formulation

Dimethyl sulfoxide (DMSO) was used as an internal standard and EOF marker (as in chapter 2). Separate samples of each acid were made up in MilliQ water and 2 μL of DMSO was added to each. The acids were also mixed at 3 different ratios with added MilliQ water and again spiked with DMSO (as per Table 3-1).

Table 3-1 Composition of the three acid mixtures to be analyzed by CE

Component	Mixture 1	Mixture 2	Mixture 3
DMSO (% (v/v))	0.118	0.012	0.012
Citric acid ($g \cdot L^{-1}$)	1	-	-
Benzoic acid ($g \cdot L^{-1}$)	0.3	0.05	0.01
2-Napthalene sulfonic acid ($g \cdot L^{-1}$)	0.1	0.1	0.1

3.2.3. Sample preparation for dissolution studies

Samples of each protein were prepared by weighing out a known amount of lyophilized powder and adding the appropriate volume of PB5 (see Chapter 2, Section 2.2.1 for buffer preparation) to achieve a concentration of ~1 $g \cdot L^{-1}$. Samples were prepared in 1.5 mL plastic Eppendorf tubes. The concentration of each was substantiated using an ultra-low volume spectrophotometer (NanoDrop, ThermoFisher Scientific) and CE experiments commenced immediately afterwards.

3.2.4. Methods

All experiments in this chapter were conducted on a Sciex P/ACE MDQ CE system[93] using oligo(sodium acrylate) in 25 mM borate buffer (pH 9.3) as a standard for the validation of the capillary and instrument. Pre- and post-treatment of the capillary was the same as in chapter 2 (Section 2.2.3). Sample injection volumes were amended from those in chapter 2 following a series of independent experiments in which different injection volumes were compared. The method used for these experiments was the same as in chapter 2 (Section 2.2.3), with only the injection volume being altered (according to Table A-3).

The individual acid samples, as well as the three mixtures were analyzed via CE and pressure mobilization (PM) following the same methods described in chapter 2 (Section 2.2.3).

For the dissolution experiments, protein samples were injected hydrodynamically at a pressure of 13.79 mbar for 8.7 s, which led to an injection volume of ~ 4.58 nL. CE was conducted at 30 kV for 7 min. Prior to each sample analysis, the capillary was rinsed with MilliQ water (5 min), 1 M NaOH (2 min) and MilliQ water again (5 min). This was followed by a deliberate 'wait' period, during which the instrument remained idle for a predetermined amount of time. A final rinse with PB5 (5 min) was implemented to condition the capillary for the subsequent sample injection. For the first 11 sample analyses in the experimental sequence, the wait time was adjusted to ensure injections were 30 min apart. Thereafter, the wait time was increased to an hour and the sample was analyzed 6 more times. Aliquots of both samples were incubated at ambient temperature for 5 days, thereafter they were analyzed again via CE.

3.3. Results and discussion

The experimental data for the evaluation of the effect of reducing the sample injection volume is presented in Table A-3. Qualitative and quantitative analysis of the results suggest that using an injection volume approximately half of what was used in chapter 2 leads to increased resolution and higher precision of dispersity values.

3.3.1. Testing a newly formulated standard

The oligo(sodium acrylate) standard uses a sodium borate buffer (pH 9.3) as the BGE due to adsorption occurring at lower pHs, including 7.4. Citric acid, benzoic acid and 2-napthalene sulfonic acid were selected for the formulation of a standard based on their ability to absorb UV and their negative charges at neutral pH. For assessing the

suitability of this acid mixture as a standard, a series of PM and CE experiments were conducted as described in section 3.2. No peak was detected for citric acid during both its individual analysis and the analysis of mixture 1 (data not shown). Citric acid was therefore excluded from further experiments. Neither benzoic acid nor 2-napthalene sulfonic acid encountered significant adsorption to the capillary wall, demonstrated by the absence of tailing in the PM data (Figure 3-2).

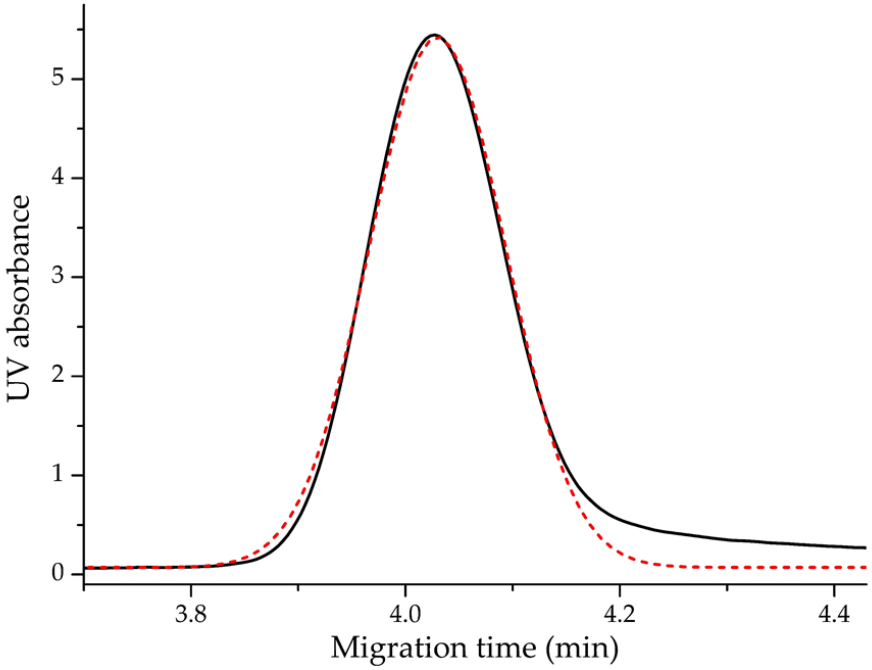

Figure 3-2 Pressure mobilization of a mixture of benzoic acid and 2-napthalene sulfonic acid (black, solid) fitted with a Gaussian function, R^2= 0.985 (red, dashed).

The electropherograms of mixtures 2 and 3 are overlaid and compared to an exemplar electropherogram for the separation of oligo(sodium acrylate) in sodium borate buffer (25 mM, pH 9.3) (Figure 3-3, a and b respectively). The mobilities of the acids in the mixture are spread across a satisfactory range to be used as a standard. An unsteady baseline following the EOF marker (DMSO peak), however, consequently leads to poor reproducibility. Due to restrictions in time, no further attempts were made in improving this formulation, and, we thus reverted to using oligo(sodium acrylate) as the standard for the subsequent dissolution experiments.

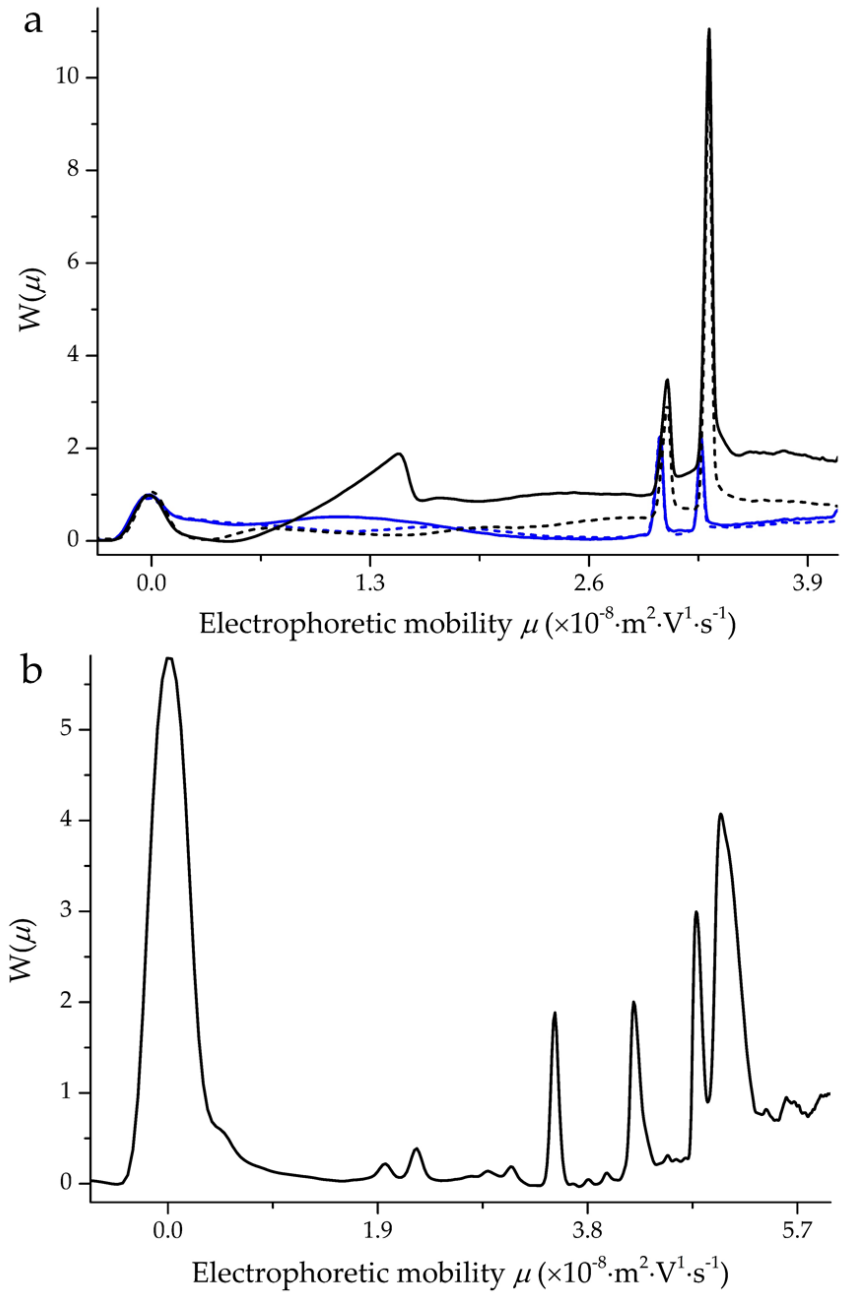

Figure 3-3 Distributions of electrophoretic mobility obtained by CE for a) mixtures 2 (black) and 3 (blue), and b) oligo(sodium acrylate). Replicates are represented by dashed lines.

3.3.2. Dissolution kinetics

The kinetics of the dissolution for each protein was monitored through sequential injections over a 15-hour time period as described in Section 3.2.4. Throughout the

duration of the experiments, samples and buffers were kept in a temperature-controlled sample tray (set to 25 °C). DMSO was used as an internal standard to partially correct for band broadening as well as variations in injection volume between each injection. The aliquots that were incubated for 5 days will be referred to as aged samples BSA and insulin. These aged samples were inverted multiple times prior to analysis via CE to ensure adequate mixing, and in the case of insulin, to resuspend the sedimented pellet that recurred upon sedimentation (Figure 3-1).

3.3.2.1. Insight into dissolution through peak integration

As stipulated by the Beer-Lambert law, a linear relationship exists between the absorbance and the concentration of the absorbing molecules in the linear range of the detector.[119] As UV detection is used in CE, the areas of the resulting peaks are proportional to concentration. Differences in peak area can therefore offer insight into the dissolution process, as an increase in area would be indicative of more molecules going into solution and vice versa. The DMSO and protein peaks were integrated on the weight-distributions of electrophoretic mobilities, $W(\mu)$, to obtain the area of each. Peak areas versus dissolution time are plotted in Figures 3-4 and 3-6.

A slight increase in protein peak area (represented by squares) is observed for both BSA (Figure 3-4) and insulin (Figure 3-6) over the first 15 hours. For both proteins, there is also an overall decrease in protein peak area for the aged samples. To determine whether these trends are accurate, the peak area of DMSO (internal standard) must also be considered (blue diamonds). The experiment was done in duplicate for BSA, with the first CE injection commencing either 15 minutes or 1.5 hours after sample preparation. The dissolution time of the first injection in the experimental sequence will be referred to throughout as $t_{initial}$.

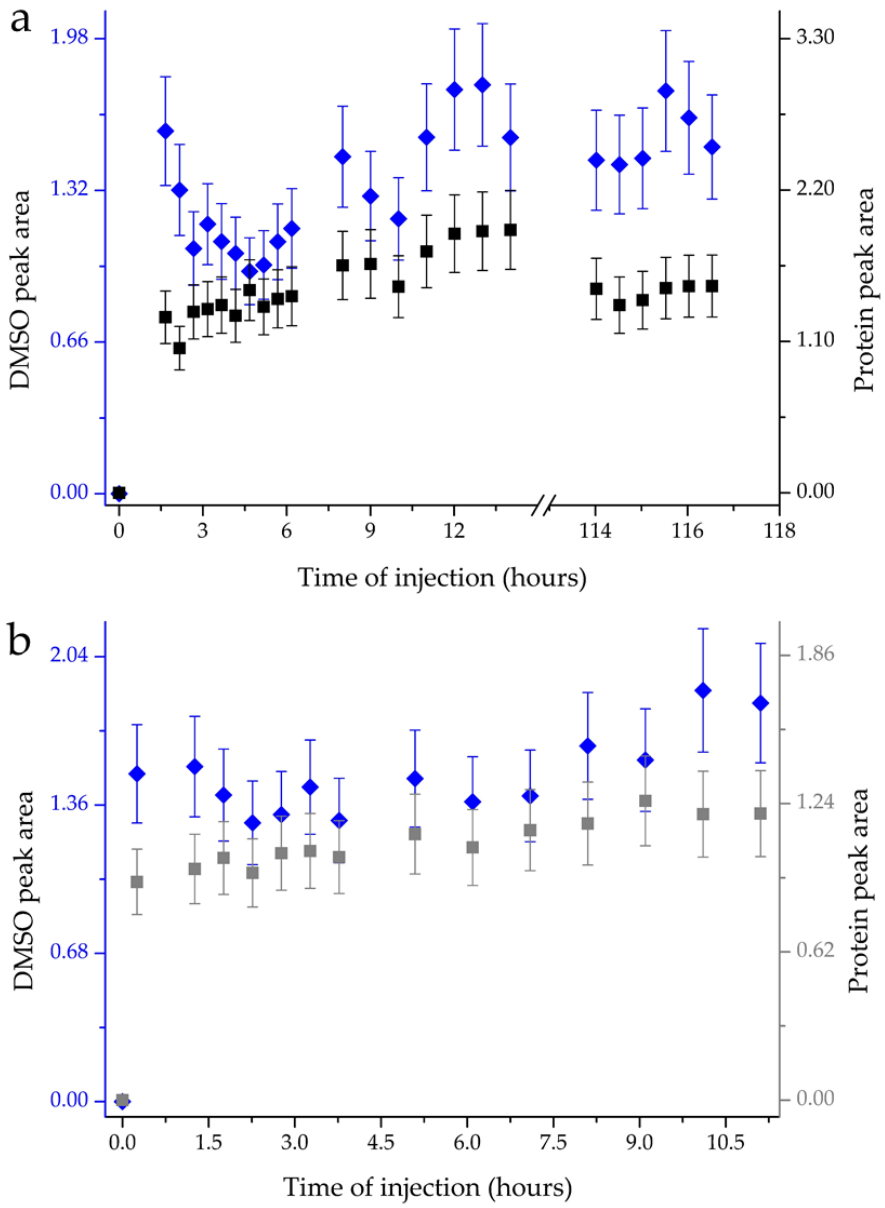

Figure 3-4 Peak area vs dissolution time of BSA (black and gray squares) and DMSO (blue diamonds) for repeat dissolution experiments with $t_{initial}$=1.5 hours (a) and $t_{initial}$= 15 min (b). Standard deviations (represented by error bars), were calculated using an RSD value of 30%.[30]

From Figure 3-4, DMSO does not follow the same linear trend as BSA during the initial ~4.5 hours. The decrease in DMSO area during this time could be the result of reduced injection volumes caused by an increase in the viscosity of the sample. Sample viscosity is suspected to increase with the concentration of dissolved protein.

The subsequent increase in DMSO peak area (Figure 3-4, a; much less prominent in b) could mean that the sample viscosity slightly decreases again as larger multimers dissociate. Alternatively, interactions between DMSO and BSA may also be influencing either the sample viscosity or the absorbance of DMSO. To account for the differences in peak area trends observed between DMSO and the protein, the data needs to be normalized. The peak area of the protein divided by the peak area of the DMSO is plotted against the dissolution time (Figure 3-5).

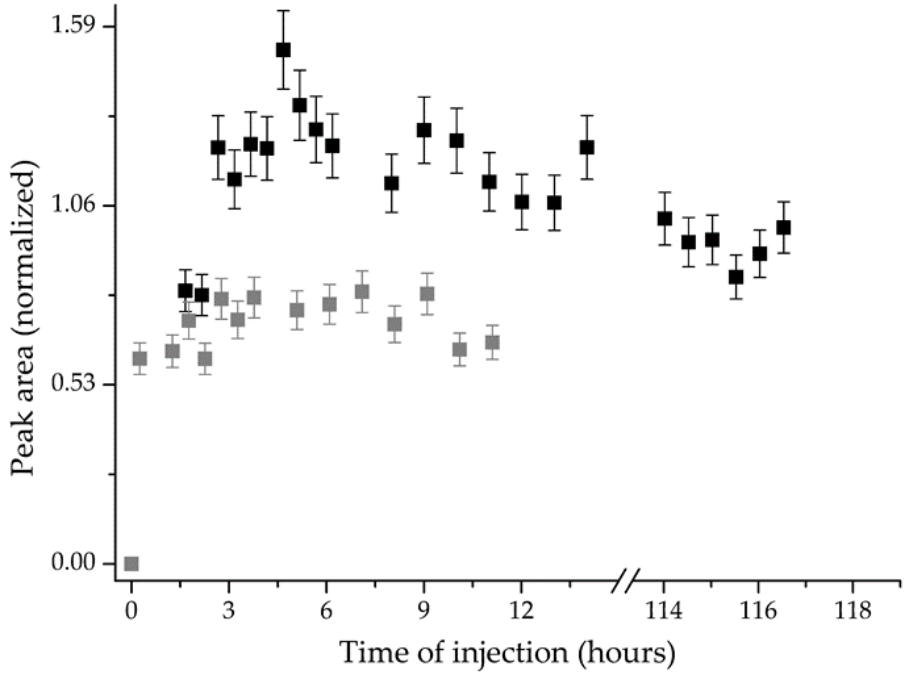

Figure 3-5 Normalized peak area of BSA peaks vs dissolution time for repeat dissolution experiments where $t_{initial}$=15 min (gray squares) and $t_{initial}$=1.5 hours (black squares) (RSD 15.29%[30]).

From Figure 3-5, the most significant increase in the normalized peak area is observed within the first 15 min of dissolution (gray data set). The normalized peak area increases by 22% and 34% from $t_{initial}$ to t=3 hours for the gray and black data sets, respectively. The following plateau likely indicates that an equilibrium in terms of dissolution has been reached within this time period. It is important to note that the definitive order of kinetics for the dissolution of BSA cannot be determined from

this data due to a lack of sufficient data points between time=0 and 15 minutes (see Figures A-1 and A-2). In summary, the results displayed as normalized peak area versus time of injection are in good agreement with the analysis of the un-normalized data and show an insignificant difference in the dissolution of BSA after 3 hours (Figure 3-5).

In the case of insulin (Figure 3-6), the first aliquot was injected 3 hours after sample preparation ($t_{initial}$=3 hours), much later than for BSA. DMSO and insulin follow the same trend (within experimental error) with regards to an increase in peak area (Figure 3-6).

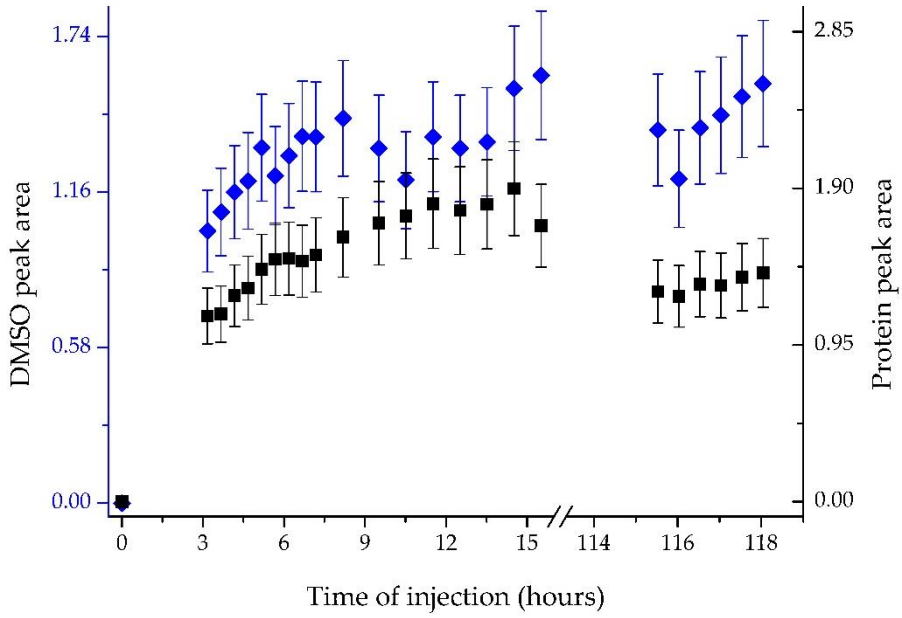

Figure 3-6 Peak area vs dissolution time of insulin (black squares) and DMSO (blue diamonds) in a dissolution experiment with $t_{initial}$=3 hours, PB5 (pH 7.4). Error bars were derived as the standard deviation calculated using an RSD of 30%.[30]

This has been attributed to a decrease in sample viscosity, resulting in larger injection volumes. The decrease in viscosity in this case could be due to dissociation of larger entities as is suspected to occur with BSA. Alternatively, it could be the result of insoluble aggregates that were initially suspended throughout the solution

sedimenting over time. The normalized peak area of insulin is plotted against dissolution time in Figure 3-7. A 13% increase in peak area at t= 9.5 hours compared to the area at t = 3 hours might be indicative of slow dissolution occurring over approximately 9 hours.

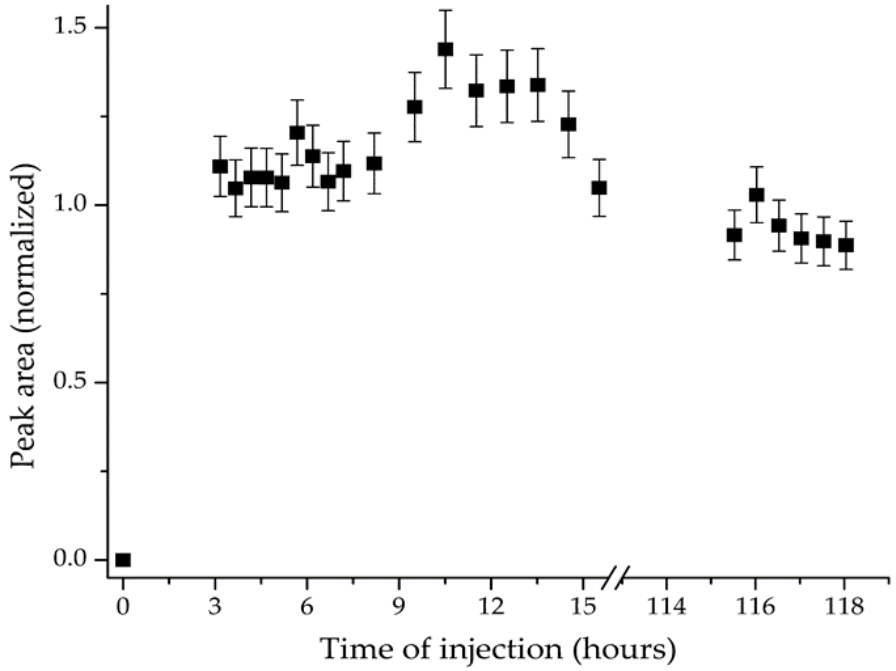

Figure 3-7 Normalized peak area vs dissolution time of insulin in a dissolution monitored from 3 hours after sample preparation (RSD= 15.29%[30]).

As initially observed in the un-normalized data, Figures 3-4 and 3-6 show an overall decrease in peak area for the aged samples (t ~115 hours). A 30% reduction in peak area for BSA and a 13% reduction for insulin is observed in comparison to the injections at 15 hours. Two hypotheses have been formulated to explain the decrease in peak area. Firstly, it is known that the UV lamp deteriorates with continuous and prolonged usage. To test this, the signal-to-noise ratio of the standard for each day of experimentation was determined and compared to one another. The differences between these ratios were not significant enough to solely explain such a large decrease in the peak area (Table A-4).

The second hypothesis is that large insoluble aggregates have either formed during incubation and precipitated out of solution (in the case of BSA) or were present in the initial suspension and sedimented (in the case of insulin). Both these scenarios would result in an overall reduction in the dissolved protein concentration. This hypothesis was derived since during incubation, the sample undergoes carbonation due to air exposure. This may decrease the pH of the solution. BSA, for one, is known to undergo various transitions of its monomer as a result of changes in the pH. These include the extended (E), normal (N) and basic (B) forms, amongst others.[50] Under acidic conditions, the extended form is adopted. Having a less compact structure means that more hydrophobic residues might be exposed to the aqueous environment. Literature therefore suggests that under acidic conditions, αshield these exposed regions and stabilize the protein structure.[50]

In order to draw definitive conclusions on the kinetics of dissolution for these two proteins, however, the method would need to be modified for the acquisition of data closer to time=0 to gain some information on what might be occuring during the early stages in which the lyophilized protein starts to go into solution (see Figures A-1 and A-2).

3.3.2.2. Assessing the conformational equilibrium through changes in electrophoretic mobility

The distributions of electrophoretic mobilities of proteins may offer insight regarding their conformational heterogeneity (see Chapter 2). Distributions were analyzed in terms of changes to the weight-average electrophoretic mobility and the heterogeneity of the sample, both during and after dissolution. It is important to note the factors that may affect both dissolution and equilibrium, inherently linking these two concepts. These include, and might not be limited to, concentration, temperature, pH and ionic strength of the solvent.[50]

Mobility data offers insight not only into the process of dissolution, but also on the physicochemical properties of the conformations present at various times. An increase in the charge to friction ratio will lead to an increase in electrophoretic mobility. As proteins multimerize and/or aggregate, regions that were previously exposed to the aqueous environment may become buried within the compact multimeric unit/aggregate. The compactness of the 'new' structure renders it less accessible for the counterion to cause frictional drag by reducing the effective charge of the overall molecule. This, in addition to a higher nominal charge, would ultimately result in an increased charge to friction ratio, thus higher mobility.[120]

The weight-average electrophoretic mobilities, μ_w, of BSA and insulin were determined from their electrophoretic mobility distributions. These values were plotted as a function of time (Figures 3-8 and 3-9) to investigate whether changes in conformation may be occurring irrespective of dissolution.

An overall decrease in weight-average mobility of BSA is observed over the first 8 hours (Figure 3-8, more prominent in the gray data set). A negative shift in the average electrophoretic mobility corresponds to an overall decrease in the charge to friction ratio of the sample molecules.

The data represented in black shows more fluctuation in the mobility of BSA; however, an overall shift towards lower mobility can still be observed. Additionally, a plateau appears to be reached after ~8 hours, indicating that a sort of equilibrium has been established. This equilibrium appears to be maintained after 5 days, evident in the unchanged mobility of the aged BSA.

Although the BSA samples used for these two repeat experiments were subjected to the exact same conditions, there are a number of factors that could influence the conformational equilibrium, yielding these slightly varying results for the two experiments. Some factors that may have influenced the equilibrium are slight

changes in protein concentration and available ions/ligands (i.e., divalent cations which may play structural roles or interactions with DMSO).

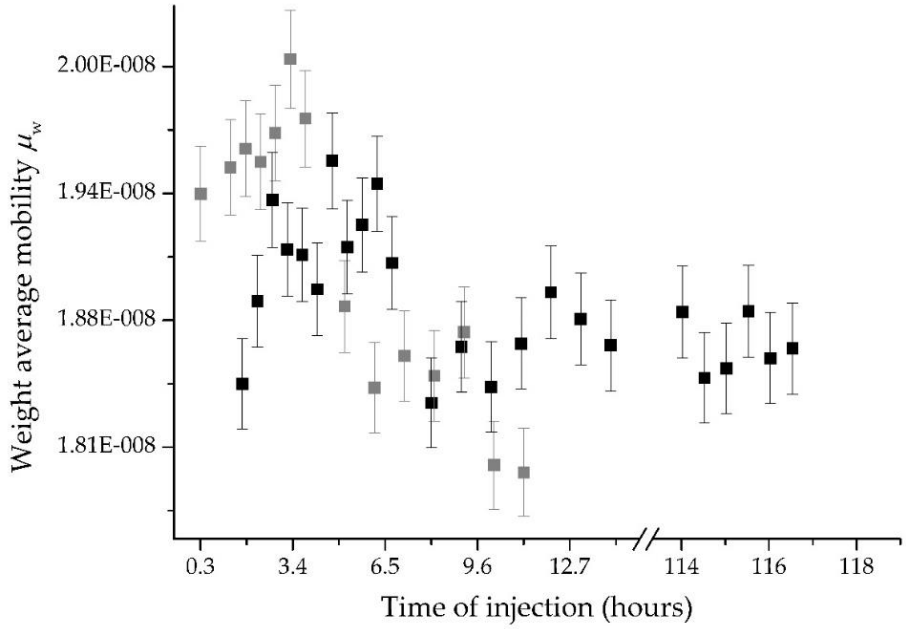

Figure 3-8 Weight-average mobility μ_w vs dissolution time of BSA in PB5 (pH 7.4) during dissolution experiments with $t_{initial}$=15 min (gray squares) and $t_{initial}$=1.5 hours. Errors are determined by the standard deviation, calculated using an RSD of 2.41% as per the literature.[92]

In Figure 3-9, an overall increase in mobility of insulin is observed, indicating a shift towards molecules with a higher charge to friction ratio. The mobility increases most notably after approximately 9.5 hours. This suggests that unlike BSA, insulin might be oligomerizing/aggregating over the course of the experiment.

A significant decrease in mobility following incubation for 5 days is observed between 15 and 115 hours. This could be explained by the theory proposed in section 3.3.3 concerning the decrease in normalized peak area, that sedimentation or precipitation of sufficiently dense aggregates has occurred, leaving only less dense/more soluble entities (with lower mobility) in solution.

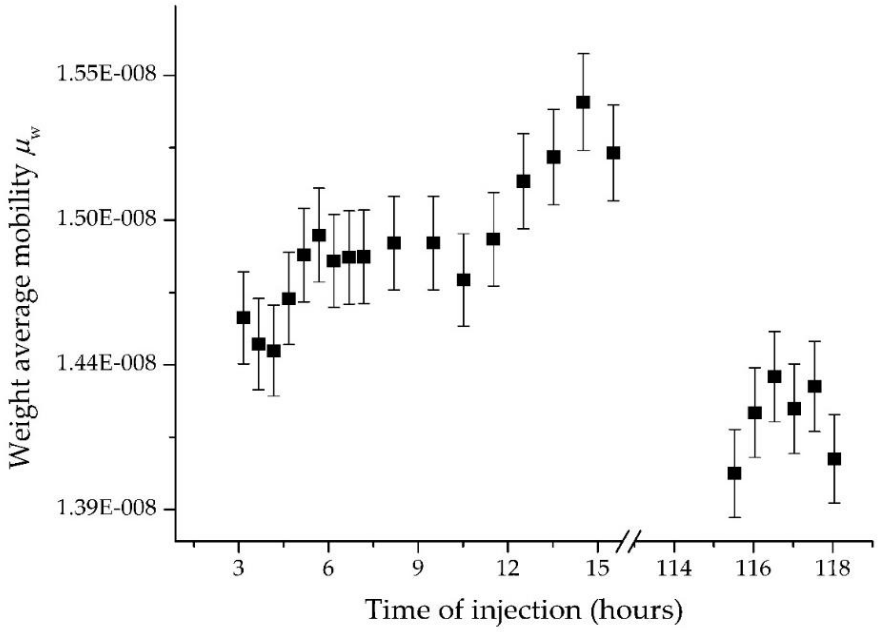

Figure 3-9 Weight-average mobility μ_w vs dissolution time of insulin in PB5 (pH 7.4) during a dissolution experiment (RSD= 2.41%).[92]

3.3.2.3. Quantification of the heterogeneity of conformations during dissolution

The weight-average mobility of a sample is useful in establishing whether conformational changes have occurred over time. Determining the heterogeneity of a sample, however, provides a quantitative measure of the variety of conformations present within the sample. On an electropherogram, peak width corresponds to the dispersion of electrophoretic mobility values. In chapter 2, dispersion is expressed as dispersity values and standard deviations (see section 2.3.3). In this chapter, we also consider the use of a more common and simple way to estimate dispersity which involves measuring the width of the peaks at half height. This is referred to as the full width at half maximum (FWHM). The measures of dispersion of the $W(\mu)$ of BSA and insulin as a function of dissolution time are presented in Figure 3-10.

For each experiment, the FWHM of both DMSO and protein was determined on the $W(\mu)$ obtained from each injection in the experimental sequence. The FWHM of

DMSO was used for the normalization of the protein FWHM to compensate for any technical factors which may lead to false trends being observed (such as band broadening, mentioned above). The normalized values of FWHM for insulin and BSA were plotted as a function of the time of injection since sample preparation (Figure 3-10, a and c, respectively).

An increase in the FWHM of insulin is observed for ~10.5 hours (Figure 3-10, a). The FWHM then decreases slightly over the following ~5 hours. A substantial increase in FWHM for aged insulin is observed, followed by decrease up to t=118 hours. The overall tendency of the FWHM for BSA (Figure 3-10, b) is an increase over time, evident in both experiments (gray and black markers). The two data sets for BSA however are not in complete agreement in terms of what is occurring within the sample at approximately overlapping times. The gray data set shows a significant increase in dispersity between 3.5 and 5 hours, followed by a nearly constant increase for the next 4 injections. The data appears to be approaching a plateau around 9 hours. The black data set shows an increase only after 6 hours, however, reaches a plateau just after 9 hours. As with insulin, a significant increase in the FWHM is observed after 5 days incubation, however unlike insulin, there is no significant change in the FWHM for subsequent times.

From Figure 3-10, similar trends are observed for the standard deviation and the $D(W(\mu),1,0)$ of insulin (c and e, respectively). Counter to the observations for the FWHM of BSA, neither the standard deviation nor the $D(W(\mu),1,0)$ follow a discernable increasing trend (Figure 3-10, d and f). The values do not show significant variation, indicating that minimal conformational changes occur within the sample, even after 5 days of incubation.

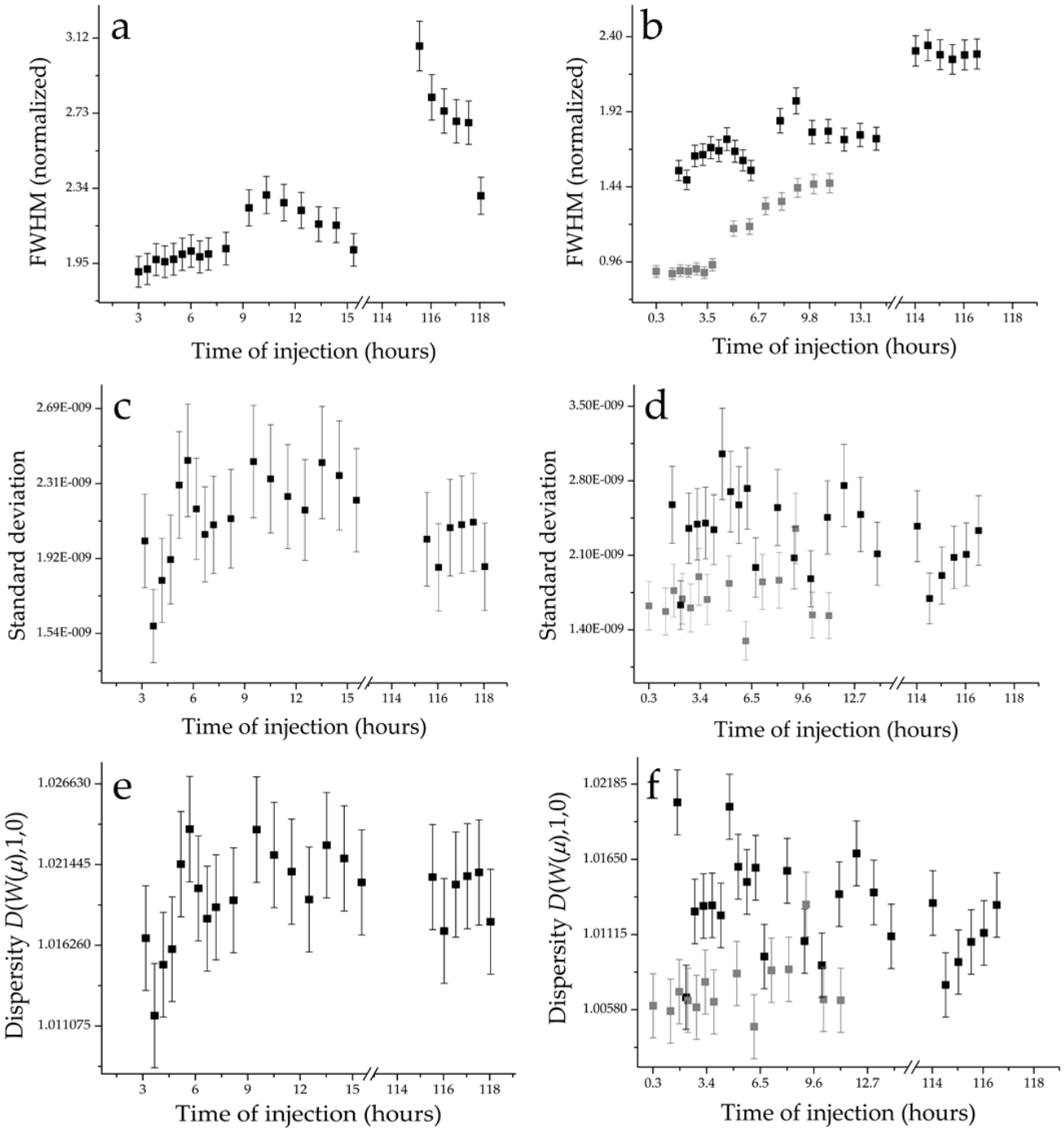

Figure 3-10 FWHM, standard deviation and dispersity plotted against dissolution time for insulin (a, c and e, respectively) and BSA (b, d and f, respectively). Grey markers (b, d and f) represent a repeat experiment. Error bars for FWHM (8.2 % RSD), standard deviation (plot c: 23 % RSD; and plot d: 28 % RSD) and dispersity (plot e: 0.68 % RSD; and plot f: 0.45 % RSD) were determined from previous experiments under analogous conditions (Table A-5).

In summary, although FWHM is a quick and straight-forward measurement and can give an indication of the sample heterogeneity, it is typically more useful as a measure of dispersion for symmetrical peaks. This is because only two points from the overall distribution are considered in calculating the FWHM. Peak shape may therefore drastically affect the measured FWHM, resulting in misleading trends which may cause false inferences regarding the sample to be made. For asymmetrical peaks, dispersities of $W(\mu)$ are more accurately expressed by the standard deviation and dispersity values obtained from the ratio of moments of electrophoretic mobility distributions (as in chapter 2).[37]

3.3.3. The implications of CE-CC in better understanding the physicochemical features of proteins

Considering all the data obtained from the dissolution kinetics of a protein known to exhibit low solubility at neutral pH, it is possible to make reasonable inferences on the conformational heterogeneity and equilibrium under given conditions. Within the first 3 hours, some of the lyophilized insulin dissolves. No further dissolution is detected over the subsequent 6 hours, as the peak area remains constant (Figure 3-7). At a neutral pH and in the presence of zinc, insulin molecules preferentially self-associate into a stable hexamers.[121] Smaller molecules (monomers and dimers) might self-associate into more stable hexamers or higher order multimers, until an equilibrium has been established.[121] The conformations that have been proposed to exist in equilibrium at neutral pH include the monomer, dimer and Zn-coordinated hexamer.[54] This, along with the allosteric nature of the hexamer enabling it to adopt at least three different conformations could contribute to an overall increase in the heterogeneity of the sample.[52] This is indeed reflected in our data by the overall increase in dispersity between 3 and 9 hours.

Over the following ~2 hours, further dissolution is apparent (Figure 3-7). Thereafter, the concentration remains constant for approximately 4 hours. The increase in

concentration either causes the equilibrium to shift in favor of the larger (more mobile) multimers, or aggregation may be occurring through any of the processes described in the literature.[54, 121] Once aggregates reach a critical size, they become insoluble and may consequently precipitate.[121, 122] This may explain the decrease in concentration at t= 14.5 hours. Despite the variations in concentration and increasing mobility, the heterogeneity remains constant (within experimental error). Since 0 is used as a reference in the calculation of $D(W(\mu),1,0)$, dispersity generally decreases as mobility increases. Therefore, depending on the degree of variance in sample heterogeneity, a shift in mobility may counteract minor differences in dispersity, rendering them negligible. This should be less apparent in the calculation of the standard deviation, however as the μ_w itself is close to 0, the same effect is likely occurring.

For the aged insulin, aggregate precipitation is suspected to have continued during the incubation period. This is once again reflected by reduced peak areas for injections at t=115-118 hours. A substantial decrease in the weight-average mobility for this data set however suggests that the sample is dominated by molecules with lower mobility. This agrees with literature in that the conformational distribution of insulin is shifted toward lower order multimers under conditions of reduced protein concentration, low ionic strength and a decrease in pH (the latter, however minimal, could be the result of carbonation).[54] The fact that there is no significant difference in the heterogeneity of aged insulin indicates that even with a reduced number of aggregates that previously contributed to a higher weight-average mobility, a proportionate number of conformations of less mobile molecules are present following incubation.

The data presented in this chapter may indicate that BSA, although known to be highly soluble at neutral pH, takes approximately 2.6 hours to completely dissolve. This is longer than the time we previously provided for BSA to dissolve prior to

experimentation (i.e., in Chapter 2). After approximately 2.6 hours, no significant difference in the peak area of BSA indicates that no further dissolution occurs over the course of the experiment. Adhering to the tentative assignment of peaks proposed in chapter 2, monomeric BSA has a lower mobility than higher oligomers and/or aggregates. It is therefore hypothesized that the negative shift in mobility observed over at least the first 8 hours (Figure 3-8) is due to the reversible transitions of higher order multimers into lower order dimers and monomers. The plateau after 8 hours indicates that an equilibrium in favor of these monomeric and dimeric species has been reached under the given conditions. The fluctuating dispersity is likely due to the high dynamism of BSA, and thus reflects the various molecular conformations it adopts during the continual process of self-association and subsequent disassociation to maintain the equilibrium.

BSA is known to form aggregates upon incubation. According to literature, aggregates of BSA are chemically different from oligomers and occur through an irreversible path involving different residues.[111] When these aggregates exceed a certain size, they may become insoluble, resulting in precipitation.[122] If we consider the marginal decrease in peak area for aged BSA, it would appear as though some aggregation has occurred during incubation, resulting in a slight decrease in concentration. No significant changes in the mobility or dispersity suggest that the conformational equilibrium is maintained during incubation.

Despite the fact that the time allowed for the sample to dissolve and equilibrate in Chapter 2 is unknown, the results are comparable to those presented here. An average mobility of 1.78×10^{-8} obtained for BSA in Chapter 2 is within range of those obtained over the course of the dissolution experiment (1.80×10^{-8} to 2.00×10^{-8}). This is also the case for the dispersity of the distribution, which in Chapter 2 averaged to be 1.007, with the range determined here being 1.005 to 1.017.

3.4. Conclusions and future work

Through this study, complexity of the relationship between dissolution and the conformational equilibrium of globular proteins was demonstrated. Insoluble aggregate formation resulting from the lyophilization process further complicates dissolution, as in the case of insulin, and prevents from obtaining a true solution. This study was able to provide an estimated amount of time required for a lyophilized sample of BSA to dissolve prior to characterization via CE. BSA is highly dynamic in its ability to reversibly self-associate and disassociate. A stable conformational equilibrium favoring the monomeric species is reached after 7 hours and maintained during incubation. Insulin undergoes slow dissolution, complicated by the temporal formation of insoluble aggregates. Aggregation leads to a constant increase in mobility, however simultaneous precipitation may mask changes in dispersity.

As the results presented here are preliminary, further assessment of the applicability of CE-CC to monitor the dissolution and conformational equilibrium of lyophilized proteins is required. This may be done by repeating experiments and optimizing some of the experimental parameters further. An important modification to be made in repeating these experiments is for CE experiments to be commenced much earlier on in the dissolution process. This could potentially enable the determination of the order of kinetics as well as the analysis of the processes occurring during early stages of dissolution. Furthermore, ^{1}H NMR can be used to monitor the dissolution of proteins and the results can be compared to those obtained by CE. This would also validate CE as a technique for dissolution monitoring. Additionally, techniques such as field flow fractionation (FFF) and Taylor dispersion analysis (TDA) could be used complementarily to CE for further analysis of the structural changes that may have occurred during dissolution.

CHAPTER 4: CE-CC as a Complementary Method to Size-based Characterization Techniques

4.1. Introduction

In this project, CE is proposed for the characterization of proteins, not to replace any existing protein characterization techniques, but to complement them for a more in-depth understanding of protein structure-function relationships. CE as a stand-alone method for the structural characterization of proteins is limited as the data produced is not sufficient in providing a clear and complete picture of the overall structure (see Chapter 2). Knowing about the electrophoretic behavior is important, however, this is only one of several properties that inform on the overall structure of a protein. It is known that the separation in CE-CC occurs due to differences in electrophoretic mobilities of the analyte molecules, and it can be assumed that these differences arise from differences in conformation. What is not exactly known, however, is what those conformational differences might be; some molecules may be more compact than others, the subunit configurations might vary slightly between molecules of the constituent sub-populations and regions of intrinsic disorder may impact the molecules behavior under an electric field. There is no one technique capable of providing all the answers regarding the relationship between the structure and function of proteins, and everything else related to it (such as misfolding and aggregation). To accommodate for these shortcomings of CE, additional techniques can be employed, either through coupling CE to them, or directly corroborating and comparing the findings with that from CE (as in chapter 2).

4.1.1. Polyacrylamide gel electrophoresis

Polyacrylamide gel electrophoresis (PAGE), as mentioned in chapter one, is one of the most commonly used techniques in proteomics. For the analysis of proteins in their native state, i.e., whether they are purely monomeric, or contain multimers or aggregates, native-PAGE can be used. In native-PAGE, sample molecules are separated by their hydrodynamic volume as well as charge.[123] Two variants of native-PAGE include blue-native PAGE (BN-PAGE) and colorless-native PAGE

(CN-PAGE). Both of these are non-reducing (i.e., disulphide bonds remain intact) and non-denaturing (i.e., no surfactants are present, and samples are not heated). They do however differ in separation principle correlated with the electrical charges on proteins.[124] In short, BN-PAGE utilizes anionic Coomassie Brilliant Blue (CBB) G-250 dye which modifies the net charge on the surface of the protein and consequently, (increases) its solubility.[124] In CN-PAGE, no CBB is added and as a result, the separation relies on the intrinsic charge of the protein instead. CN-PAGE is milder then BN-PAGE but faces some limitations such as decreased resolution and being restricted to acidic proteins.

Native-PAGE remains a useful tool for obtaining insight into the sizes of the molecules present in a single sample, which is related to both molar mass and conformation. A two-dimensional separation employing native-PAGE in the first dimension and CE in the second might allow for a more comprehensive investigation into the conformational changes and protein-protein interactions that might occur in a sample.

4.1.2. Taylor dispersion analysis

So far, the work presented has been focused on the potential of CE to separate proteins based on conformational differences and the ability to monitor adsorption through complementary pressure mobilization experiments. When adsorption is negligible, another useful feature of the CE instrument itself is that it can be used to determine the weight-average hydrodynamic radius of an analyte through Taylor dispersion analysis (TDA).[125] Similar to pressure mobilization, TDA involves injecting a solute plug and applying hydrodynamic pressure to drive the sample through the capillary.[126] The data is presented in the form of a Taylorgram where absorbance is plotted as a function of time.[127] The broadness of a peak is quantified via the temporal variance of the elution profile, which is calculated by fitting a Gaussian function to the peak.[125, 128, 129] This experiment is done at various pressures

and the values obtained can in turn be used to determine an average molecular diffusion coefficient (D) as well as an average hydrodynamic radius of the analyte species in the sample.[126, 130]

One important advantage of TDA compared to other macromolecular sizing techniques, such as the aforementioned PAGE techniques, is the absence of a stationary phase allowing for a batch analysis in free solution.[126] Other advantages are parallel to those for CE, namely decreased volumes of sample required and waste produced.[129] On the other hand, TDA does not provide size-distributions. TDA produces an average value of hydrodynamic radius, where CE-TDA or SEC-MALS still holds the advantage of providing distributions of molar masses/sizes within the sample. This is often more informative, depending on the nature of the sample and what the data is needed for. As monitoring and minimizing adsorption of the sample to the capillary wall is crucial before attempting TDA, the focus of pressure driven propagation experiments implemented in this thesis is focused on adsorption analysis, rather than for sizing purposes.

Due to the aforementioned limitations of TDA, another group of techniques, known collectively as field flow fractionation (FFF) was explored. FFF overcomes some of the most common drawbacks faced in SEC and TDA. As this technique has already been briefly introduced in chapter 1, the following section will focus on a specific subclass, known as asymmetric flow field flow fractionation, or simply AF4.

4.1.3. Asymmetric flow field flow fractionation

AF4 is the most commonly and successfully used FFF technique for protein separation.[131] The sample is introduced into a flat channel (~100-500 μm in height) by a carrier flow. This creates a parabolic flow profile, with the fluid in the center of the channel having the highest velocity. The lower wall of the chamber consists of a semi-permeable membrane which allows for a cross flow to be introduced perpendicular to the carrier flow. This crossflow can either be pressure-driven or

fluid and drives the sample to the semi-permeable membrane, known as the accumulation wall (Figure 4-1).[131, 132]

The crossflow drives the sample components towards the wall at a given velocity *U*, where they are semi-stagnant due to the parabolic flow profile. At the same time, the components also undergo diffusion away from the wall. Their mean distance from the accumulation wall is expressed as a ratio of *D*/*U*.[38] In AF4, *U* is constant for all sample components, and thus solute retention is reliant on the diffusion coefficients *D* of the components, which is related to their hydrodynamic radii through the Stokes-Einstein equation.[131, 132] Small analytes with larger *D* will elute first followed by larger analytes with a smaller *D*. The lower cut-off for the analytes is generally determined by the cut-off of the membrane used, typically 1,000 to 30 ,000 kg·mol^{-1}.

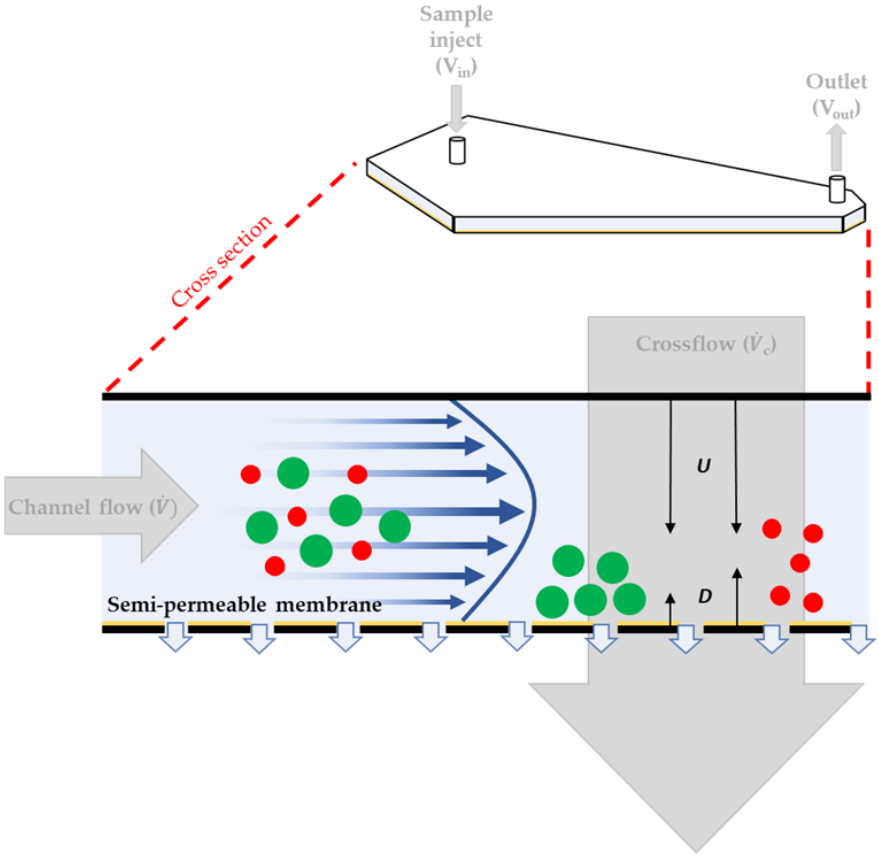

Figure 4-1 A cross-sectional view of an AF4 channel showing the basic separation mechanism. Macromolecules with a smaller hydrodynamic volume have an increased *D*, which drives them into the higher flow rate regions of the parabolic flow profile, leading them to be separated out first. Image inspired by literature.[133]

The retention data obtained experimentally from FFF experiments can be used for the subsequent determination of important properties of the various components within a sample, such as molar mass.[131] The diffusion coefficient can be calculated using an equation which incorporates a number of experimental and derived variables/parameters.[132] The following equation can be derived for well retained analytes, i.e., when the ratio of the void peak retention time (t^o) and the analyte retention time (t_r), defined as the retention ratio (R), is less than 0.3:[134]

$$D = \frac{t^o \dot{V}_c w^2}{6 V^o t_r}$$

where w is the channel width, $\dot{V}_c$ is the flow rate of the perpendicular crossflow and V^o is the void volume (which are all experimental parameters).

Multi-angle light scattering (MALS) detectors measure the intensity of scattered light at various angles simultaneously. In the case that the wavelength of the applied light is longer than the physical dimensions of the analyte particle, Rayleigh scattering will occur.[135] In such cases, the average molar masses of the components within a given sample can be determined through the use of the expression of Rayleigh ratio given below.[136]

$$I_{scattered}(\theta) \propto R(\theta) = K^* M c P(\theta)[1 - 2A_2 M c P(\theta)]$$

Where $I_{scattered}(\theta)$ is the intensity of the scaterred light at a given angle θ, $R(\theta)$ is the Rayleigh ratio (from the solute), K^* is the contrast factor, M is the molar mass in g·mol^{-1}, c is the solute concentration in g·mL^{-1}, $P(\theta)$ is the particle form factor and A is the second virial coefficient.[136]

The data obtained by AF4 can also be used to calculate the dispersity of molar mass distributions by employing the same calculation as is used for this purpose in the treatment of SEC data.

4.2. Experimental section

Lyophilized samples of BSA and α-LA were from the same batches as those used in Chapters 2 and 3. Additional lyophilized samples of ADH, catalase (from bovine liver) and deoxyribonuclease I (DNAse I, from bovine pancreas) were purchased from Sigma-Aldrich. The concentrations of the protein samples varied between experiments (ranging from 0.5 to 4.5 $g \cdot L^{-1}$), however, were determined in the same way using their respective extinction coefficients ($E^{1\%}$), as listed in Table A-1. Concentrations were measured spectrophotometrically approximately 10 min after adding PB5 to the lyophilized protein powders (as in Chapter 2). The water used throughout all experiments was of MilliQ standard (18 MΩ cm-1, 0.22 μm membrane filter). All other reagents used in each of the following experiments were purchased from Sigma-Aldrich, unless specified otherwise.

4.2.1. Colorless-native polyacrylamide gel electrophoresis (CN-PAGE)

4.2.1.1. Preparation of SDS-free polyacrylamide gels

Stock solutions of 0.5 M (pH 6.8) and 1.5 M (pH 8.8) Tris buffer were prepared by weighing out appropriate amounts of Tris base, adding it to the required volume of water and titrating it with HCl. These buffers were used for the subsequent preparation of colorless native polyacrylamide gels (60 × 70 × 1 mm in size). The gels were hand cast at ambient temperature and consisted of a 1% stacking gel (26.79% (v/v) 40% acrylamide/bis (37.5:1), 47.63% (v/v) MilliQ, 24.81% (v/v) 1.5 M Tris, 0.70% (v/v) 10% ammonium persulfate (APS), and 0.07% (v/v) Tetramethyl ethylenediamine (TEMED)) and a 10% running gel (8.95% (v/v) 40% acrylamide/bis (37.5:1), 65.61% (v/v) MilliQ, 24.85% (v/v) 0.5 M Tris, 0.50% (v/v) 10% APS, and 0.10% (v/v) TEMED).

4.2.1.2. Sample preparation

A stock (4X) sample buffer with the following composition was prepared and stored as 500 µL aliquots: 20% (v/v) 1 M Tris-HCl (pH 6.8), 0.4% (w/v) bromophenol blue, 32% (v/v) glycerol (ThermoFisher), and 48% (v/v) MilliQ. Samples of protein (ADH) were prepared in 1X sample buffer at concentrations ranging from 0.72 to 3.6 $g \cdot L^{-1}$.

4.2.1.3. Method

Electrophoresis was performed at ambient temperature using a Bio-Rad Mini-PROTEAN Tetra cell coupled to a Bio-Rad PowerPac Basic Power Supply. Approximately 20 µL of protein sample (per well) was loaded onto the gel for electrophoresis alongside a pre-stained standard (Precision Plus Protein Dual Color standard, 10-250 kDa, Bio-Rad). The running buffer was a tris/glycine buffer composed of 1.44% (w/v) glycine and 0.303% (w/v) tris base in water. The voltage was set to 120 V, and the experiment was left to run for approximately 45 min. After electrophoresis, gels were either stained for 2 hours with 1× R-250 CBB (40% methanol, 7% acetic acid, 0.125% CBB), and de-stained in water overnight, or alternatively stored in running buffer overnight. Three different sample concentrations were first compared. A concentration of 3 $g \cdot L^{-1}$ yielded the best results in terms of resolution and band intensity according to visual assessment. After imaging (Gel Doc EZ Imager, Bio-Rad) CBB stained gels, elution was performed (Mini Whole Gel Eluter, Bio-Rad), under a current of 75 mA for 30 min. The fractions were manually collected for further analysis by CE.

4.2.2. Pressure mobilization (PM)

All experimental parameters and conditions were identical to those established in chapter 2 (Section 2.2). Samples of catalase (1.14 $g \cdot L^{-1}$) and DNAse I (0.75 $g \cdot L^{-1}$) were prepared and analyzed on the same day as those in chapter 2.

4.2.3. Asymmetric flow field flow fractionation (AF4)

4.2.3.1. Sample preparation

An 8.5 mM sodium phosphate buffer (PB 8.5) with the following composition was prepared: 0.071% (w/v) Na_2HPO_4, 0.059% (w/v) NaH_2PO_4, 0.88% (w/v) NaCl, and 0.02% (w/v) sodium azide.[105] The buffer pH was adjusted to 7.4 using 5.0 M NaOH, after which it was sonicated for 10 min and filtered. An aliquot of the buffer was used for the subsequent preparation of samples of α-LA (2.15 $g \cdot L^{-1}$) and BSA (1.70 $g \cdot L^{-1}$). The remaining buffer was used as the carrier fluid (mobile phase) for the experiments.

4.2.3.2. Refractive index increment determination

The refractive indices (RI) for α-LA and BSA were measured at ambient temperature using a digital refractometer (SPER Scientific, model no. 300034) equipped with a sodium light source (620 nm). The concentration series consisted of at least 4 points for each protein in the concentration range of 0.5 $g \cdot L^{-1}$ – 4.5 $g \cdot L^{-1}$. The data was used to construct a calibration curve (Figure 4-6) in order to determine the refractive index increment (dn/dc) for each respective protein (as in chapter 2).

4.2.3.3. Instrumental setup

AF4 experiments were performed on an Agilent 1200 HPLC system (Agilent Technologies) constituting a degasser (G1322A), quaternary pump (G1311A), autosampler (G1329A) and a 280 nm UV detector (G1315) .This was coupled to a Wyatt Dawn Heleos II multi-angle light scattering (MALS) detector with inbuilt quasi electric light scattering (QELS) and refractive index (RI) detection, as well as a Wyatt Eclipse 3+ A4F system. The membrane used for both proteins was a 10 kDa regenerated cellulose membrane (Superon) and the spacer size was 350 μm. The software used to set the AF4 separation parameters was ChemStation (version

B.04.02) (Agilent Technologies). The data was acquired using ASTRA software (version 6.1) and treated using both ASTRA and Origin (version 9.0) software.

4.2.3.4. Method

The method for a single sample separation consisted of a preliminary rinse of the membrane which involved the carrier fluid (PB 8.5) being passed through the channel at a flow rate of 1 mL·min^{-1} for a duration of 2 min. Thereafter, a focusing flow (V_{focus}) with a flow rate of 1.5 mL·min^{-1} was introduced in the opposite direction to the longitudinal detector flow. This initial focusing was maintained for 1 min before 20 µL of sample was injected at a flow rate of 0.2 mL·min^{-1} via an additional/separate injection inlet for 1 min (V_{inject}), which was followed by another 5 min of focusing. The V_{focus} was then switched off, and a crossflow (V_{cross}) of 3 mL·min^{-1} was introduced. After 15 min, the V_{cross} was switched off, leaving only the detector flow, which allowed the remaining sample to elute from the channel. Between each new sample/ experiment, the channel and pipes were rinsed by flushing the carrier fluid through the system at a flow rate of 1 mL·min^{-1}, zero cross flow and 0.2 mL·min^{-1} inject flow.

4.3. Results and Discussion

4.3.1. CN-PAGE

Due to CN-PAGE being the mildest form of PAGE, it offers the best chance to analyze physiological supramolecular structures. The results (Figure 4-2) were used to assist with the interpretation of the CE data obtained in chapter 2.

Three protein bands were detected for ADH which visually correspond to molar masses of approximately 130 000, 70 000 and 50 000 g·mol^{-1}. A substantial amount of smearing is present in all three replicates, and, to a lesser extent, in the standard. As stated previously (section 4.1.1), native-PAGE separates analytes according to hydrodynamic volume as well as surface charge. The influence of surface charges on

proteins' separation thus often leads to smeared bands such as those observed here, which results in inaccurate molar mass determinations.[137] Several modifications of the method have been proposed to address these issues[137], however for the purpose of this study, we attempted to avoid any modifications that might alter the stability and/or structure of the protein. Instead, PAGE was used to identify the number of populations present and determine their multimeric states though the approximate estimation of their molar masses. The retention factors (R_f) of the bands were determined using plots of band intensity versus R_f (Figure 4-3) generated by the imaging software (Bio-Rad Image Lab 5.2.1). The retention data of the protein standard (Figure 4-3, a) was used to construct a standard curve to estimate the molar masses of the ADH sub-populations (Figure 4-4).

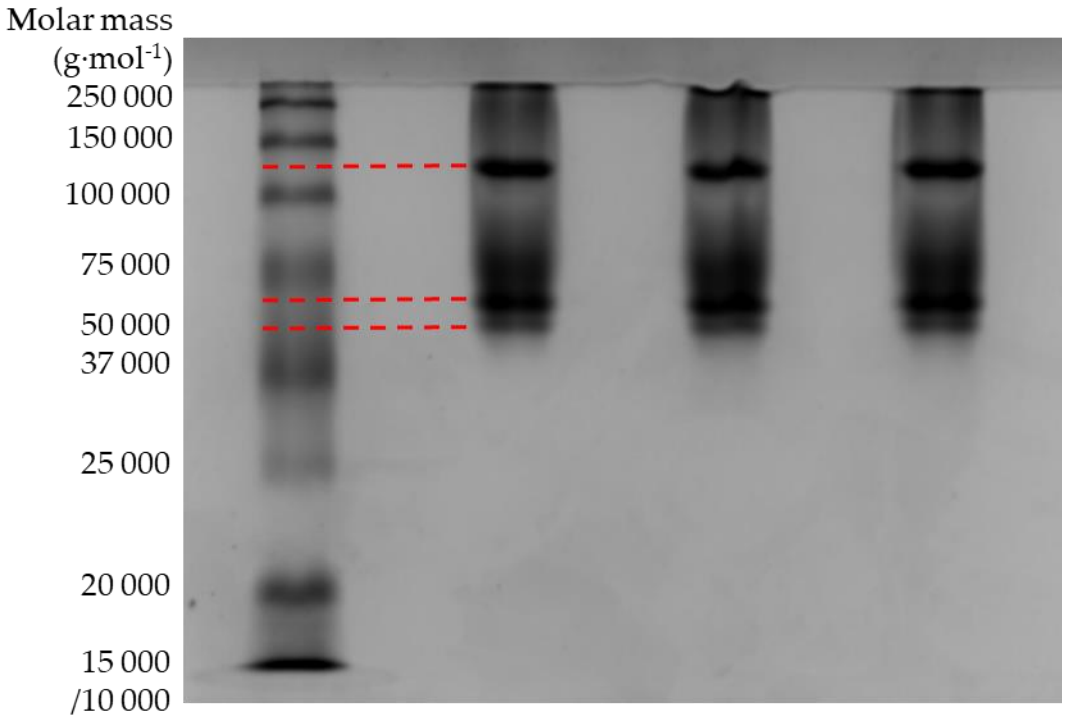

Figure 4-2 CN-PAGE of 3 g·L^{-1} ADH (in triplicate) using a Precision Plus Protein standard from Bio-Rad (far left lane) in a 10% polyacrylamide gel, pH 7.4. The red dashed lines are a guide to the eye to indicate the position of the three bands observed for ADH.

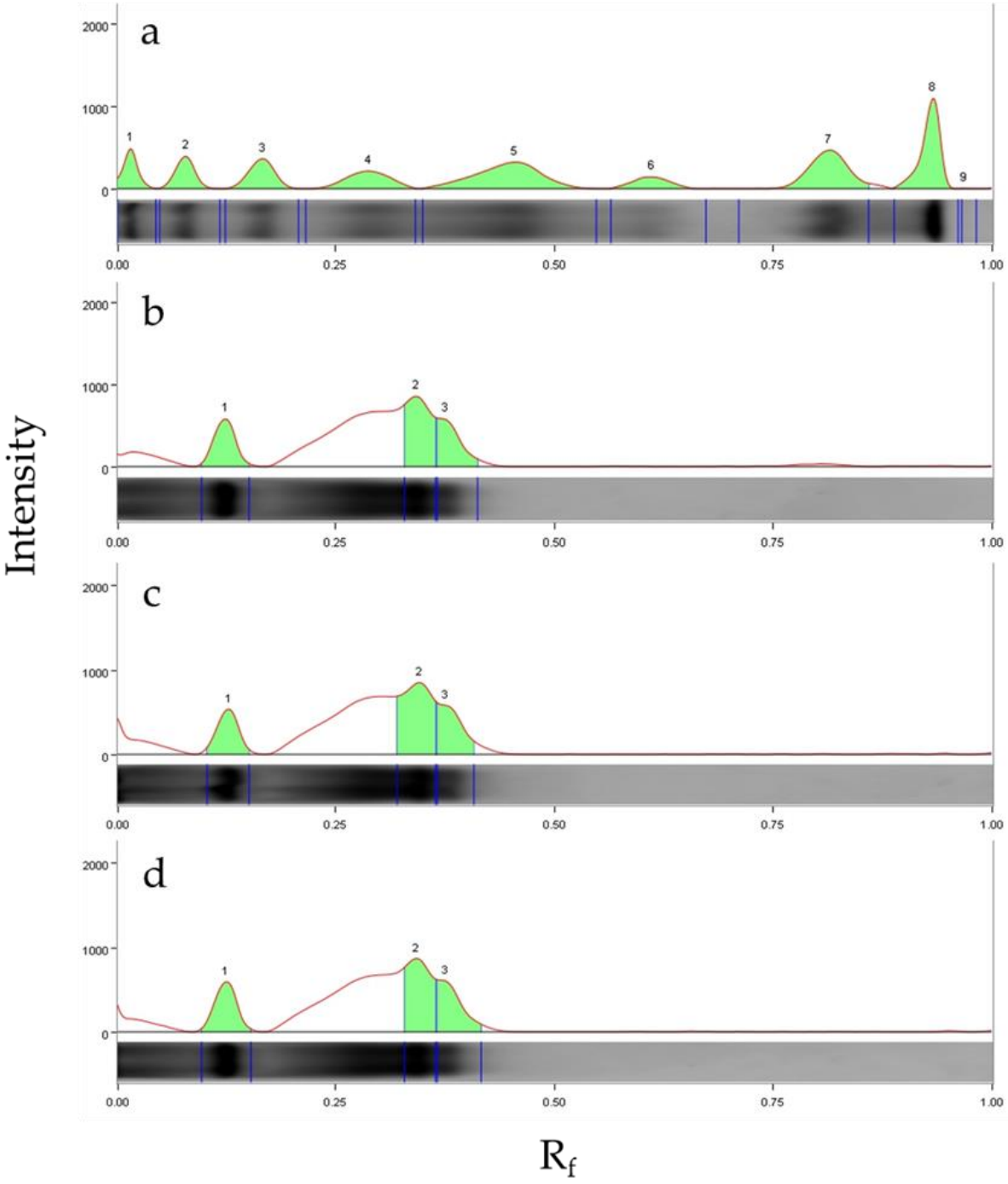

Figure 4-3 Lane profiles generated by Image Lab (5.2.1) indicating the band intensity plotted against R_f. Green shaded areas correspond to visually distinct bands detected on the gel. Panel a shows the separation of the Precision Plus Protein standard, and panels b, c and d show replicated separations of 3 $g \cdot L^{-1}$ ADH.

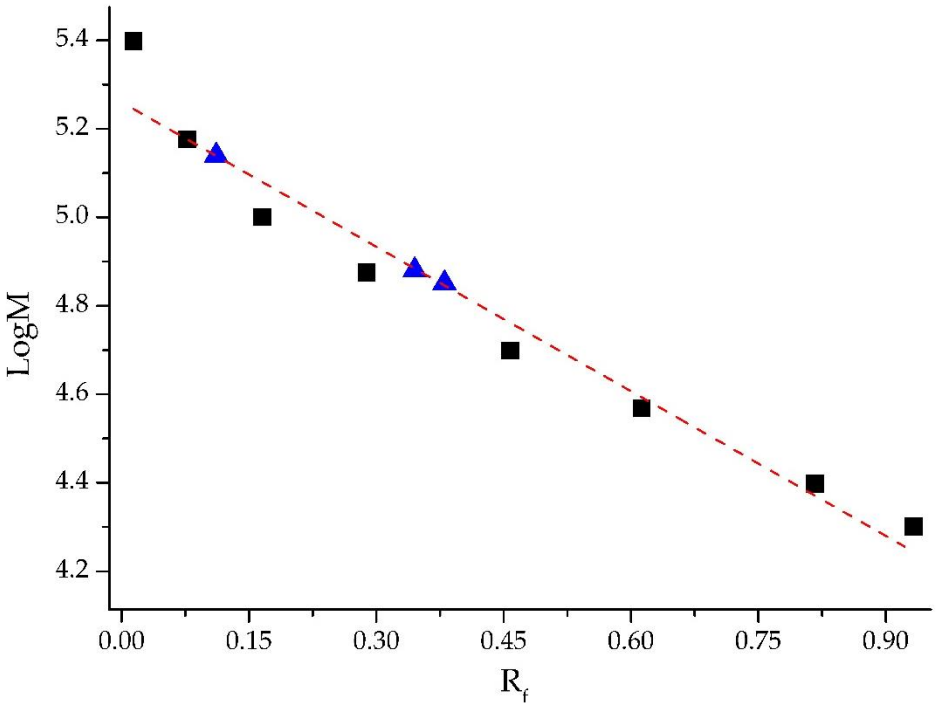

Figure 4-4 Log(Molar mass) versus R_f plotted using the Precision Plus Protein standard (black squares), $R^2 = 0.95$. The equation of the line ($y = -1.089x + 2.260$) was used to estimate the molar masses of ADH populations (blue triangles).

R_f values were determined from Figure 4-3. A band with an R_f value (0.111) corresponding to a molar mass of approximately 138 000 g·mol^{-1} is suspected to be the tetrameric form of the protein, which has a known molar mass between 141 000-151 000 g·mol^{-1}.[39, 106, 138] The monomeric and dimeric forms of ADH are therefore predicted to have molar masses ranging from 35 000 to38 000 g·mol^{-1} and 71 000 to 76 000 g·mol^{-1}, respectively, as this protein is classified as a homo-tetramer. In contrast to the SEC results for this protein (Chapter 2) where only two populations are apparent corresponding to the monomeric and tetrameric forms (Figure 2-4, c), the additional two bands detected by PAGE are visually more in-line with the dimeric form. Their calculated R_f values were 0.345 and 0.378, which correspond to approximate molar masses of 76 000 g·mol^{-1} and 71 000 g·mol^{-1}, respectively. The presence of two bands is suspected to result from dimers of slightly different conformations.[39] Four crystallographically different subunits with similar structures

which are arranged as two dimers (referred to as AB and CD) has been described in the literature.[39] The apparent band at R_f = 0.281 has been attributed to smearing, also evident in Figure 4-4.

The different populations detected in CN-PAGE and SEC could be due to different solvents being used for each experiment. The specific interactions of the protein with the constituent solvent molecules would however need to be further investigated to definitively conclude whether this might be the case. The presence of multiple conformations in either case is supported by the CE results of ADH (see Chapter 2, Figure 2-4, a and b), where a broad distribution of electrophoretic mobilities with discernable shouldering is evident. This indicates high conformational dispersity. It may also be worth noting that ADH had the highest dispersity with regards to its electrophoretic mobility distribution of all three proteins studied in Chapter 2 (Table 2-1).

To further analyze the sub-populations of ADH, a two-dimensional approach was also attempted in which the proteins were eluted from the gel, the fractions collected and subsequently injected into the CE for analysis. As described in the methods (Section 4.2.1.3), two gels were run simultaneously. One was stained with CBB for band visualization, while the other remained stain-free for elution. This was done so that the fractions collected after eluting the protein from the gel would not contain CBB which would influence the subsequent CE analysis. A numbered cutting template was placed on the stained gel to more precisely predict where the corresponding bands should be located on the unstained gel. The numbers on the template coincide with the fraction-collecting wells of the eluter and thus can be used to indicate which fractions should contain sample. The CE results of these fractions did not reveal any peaks that could have corresponded to the protein (Figure A-4, a). Two possible explanations for this were considered, both relating to the concentration of protein in the collected fractions. The first being a question of

whether sufficient amounts of protein are recovered using the elution apparatus. PAGE gels are known to show strong affinity for proteins[16], making subsequent sample retrieval inherently challenging. Secondly, separation in the first dimension inadvertently dilutes the sample. This could have led to the protein concentration in the collected fractions being too low to be detected in CE.

To investigate this, a repeat experiment was conducted in which the protein was eluted from a CBB-stained gel and analyzed by CE. As CBB binds to protein, visual observations of the collected fractions could indicate whether they contain protein at all. Additionally, CBB absorbs UV well, thus could increase the sensitivity of detection in CE for samples of low concentration. CE of these fractions however still revealed no peaks corresponding to the protein (Figure A-4, b). This could be due to a number of reasons. One explanation was that the CBB stain increases the net negative surface charge of the protein upon non-covalent binding of the dye molecules to the protein, causing the protein to have a higher electrophoretic mobility. This charge-shift forms the basis of the BN-PAGE method (see section 4.1.1). As the structure of CBB R-250 (used here) differs to that of CBB G-250 (used in BN-PAGE) only in a pair of methyl groups attached to the triphenylmethane skeleton[139], the effect on the overall charge of proteins was presumed to be similar. To test this, repeat CE experiments were performed in which the time allowed for separation was tripled. The results revealed no peaks for the protein (Figure A-4, b).

4.3.2. PM of adsorbing proteins

As suggested in section 4.3.1, proteins adsorption to the stationary phase during separation leads to inaccurate molar mass or electrophoretic mobility determination. It also makes their recovery for further analysis rather challenging. In CE, adsorption is minimized by the absence of a stationary phase and, as mentioned in Section 2.3.1 (Section x, p. x), any adsorption to the capillary wall can be detected by analyzing the temporal propagation pattern produced by pressure mobilization experiments

through comparison of the area and symmetry to that of a non-adsorptive control.[140] In the case where the proteins experience a non-negligible degree of adsorption, experimental alterations need to be made to ensure the data accurately represents the entire sample. In Figure 4-5, the PM results of two adsorbing proteins are presented. The peaks were fitted with a Gaussian function to assess how well the data is fitted to the regression line. This gives an indication of the degree of adsorption. The R^2 values were determined as 0.966 and 0.969 for catalase and DNAse I, respectively. These values are lower than those obtained for the proteins analyzed in Chapter 2 (Table A-2).

At pH 7.4, the silanol groups on the inner wall of the capillary are negatively charged. As with the proteins studied in Chapters 2 and 3, the pIs of the proteins analyzed here are lower than the pH. Therefore, they bare a considerably negative charge under the given conditions, which promotes Coulombic repulsion which is expected to minimize adsoption.[140]

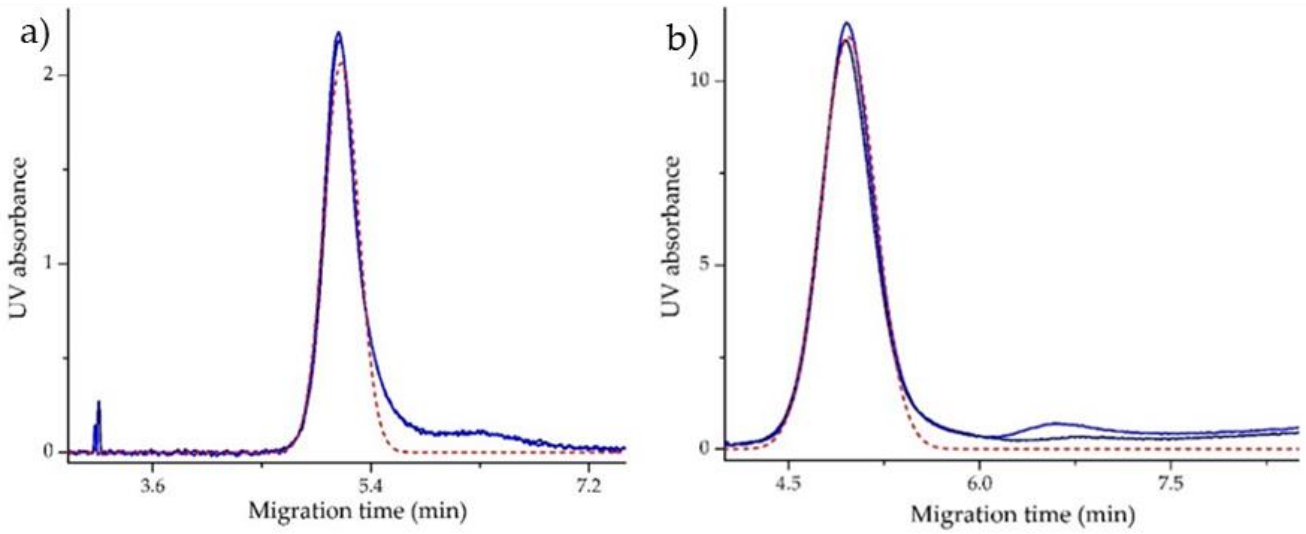

Figure 4-5 Pressure mobilization of a) DNAse I and b) catalase at pH 7.4, propagated under 50 mbar pressure.

Due to proteins being amphiphilic molecules, however, electrostatic attraction to the capillary wall could still occur due to the locally contained regions bearing a net

positive charge.[140] Additionally, depending on how they are folded, the positive regions of catalase and DNAse I may be more exposed than those of the proteins from Chapter 2. Further investigation is required to support this theory however this extends beyond the scope of this study.

Furthermore, a model for the characterization of protein adsorption onto capillary walls is described in the literature.[102] Various propagation profiles related to adsorption have been simulated. These profiles are dependent on three characteristic times, namely the time of adsorption (t_{ad}); time of desorption (t_{de}) and time of detection (t_{det}).[102] A qualitative comparison between these simulations and the results presented here was carried out. The propagation patterns for both DNAse I and catalase (Figure 4-5) appear to be congruent with that of a sample which undergoes fast adsorption to the capillary wall, and slow desorption when exposed to sample-free buffer (that is, $t_{ad} < t_{de}$). When $t_{de} \sim t_{det}$, a characteristic "bump" can be observed. Two principal ways in which adsorption can be overcome in CE are altering the pH and composition of the BGE and/or applying a coating to the inner capillary wall. As maintaining physiological pH is of priority for the purpose of this thesis, a more viable option would be to implement coated capillaries for the characterization of catalase and DNAse I by CE.

4.3.3. AF4

For the determination of molar mass from the intensity of scattered light, one essential parameter that needs to be known is the experimental refractive index increment of the analyte.[141] The refractive index increment, or dn/dc, can be concisely defined as the change in the refractive index of the solution as a function of protein concentration.[142] Although unmodified proteins have been demonstrated to have little variation in their dn/dc values, applying a consensus value for all proteins may be inaccurate as a result of varying charge effects and ion binding.[141] This is relayed in the discrepancies between the experimentally measured dn/dc values for

BSA from the literature. A number of literature values have been compiled in Table A-6, showing the variance in dn/dc for the same protein under different experimental conditions. As a result of the difficulty in selecting a value from the literature for conditions analogous to ours (Section 4.2.3), the dn/dc values for α-LA and BSA were determined experimentally under the appropriate conditions as described in Section 4.2.3.2.

From the calibration curves constructed as per Section 4.2.3.2, the slope of the regression line fitted to the refractometry data for each protein was taken as its respective dn/dc (Figure 4-6). The dn/dc values obtained were 1.90 and 2.17 for BSA and α-LA, respectively. These values are in good agreement with literature values and were therefore used for the subsequent molar mass calculations.

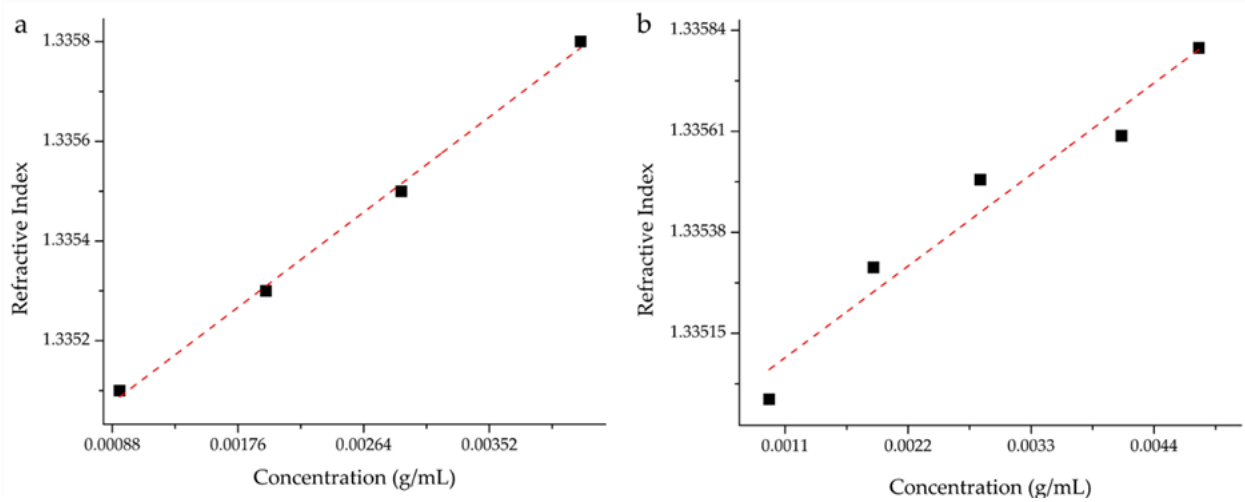

Figure 4-6 Calibration curves of the RI versus concentration (g·mL^{-1}) of (a) α-LA (R^2 = 0.996) and (b) BSA (R^2 = 0.938) in PB 8.5 + 100 mM NaCl (pH 7.4) at ambient temperature.

AF4 experiments were performed on two of the proteins analyzed in Chapter 2 in order to obtain distributions of molar masses (Figure 4-7). The distribution of molar masses for α-LA obtained by AF4 (Figure 4-7, a) is in good visual agreement with the SEC results presented in Chapter 2 (Figure 2-3) in that a single, monomodal peak is observed for both techniques.

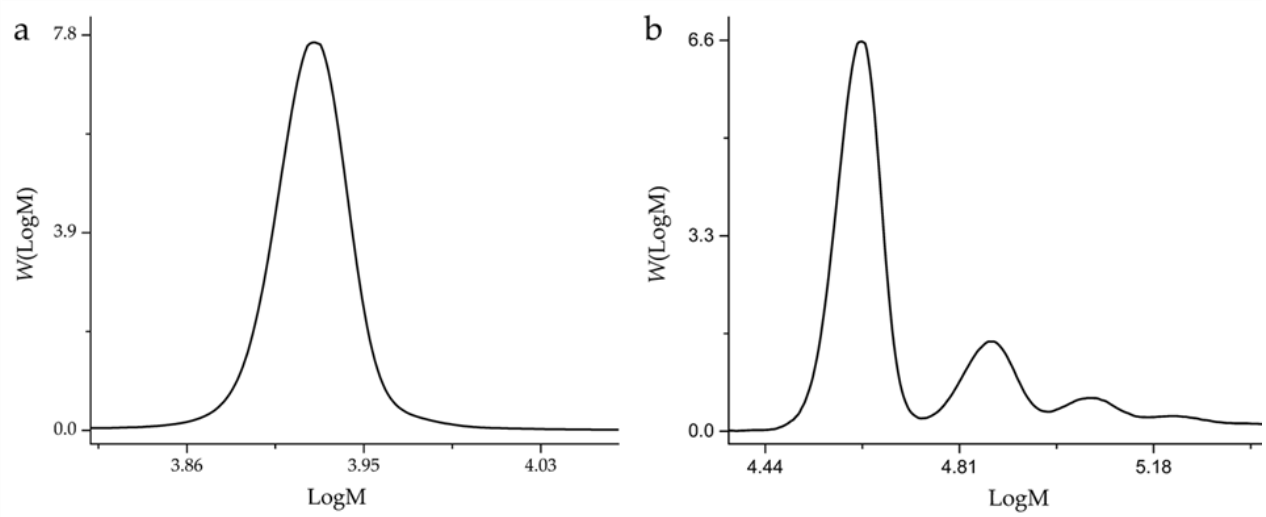

Figure 4-7 Molar mass distributions of a) α-LA and b) BSA calculated from UV detection and light scattering, after AF4 separation using a regenerated cellulose membrane with a cut-off of 10 000 g·mol^{-1}

The recovery was estimated by the ASTRA software as 89%. AF4 of BSA yielded a recovery of 74%. The molar mass distribution (Figure 4-7, b) obtained for BSA was visually similar to the one obtained by SEC, however higher oligomers were also observed in the AF4 data. The lack of these higher oligomers in the distribution obtained by SEC strongly indicates that shear degradation or filtering may have occurred as the BSA molecules migrated through the SEC column.[133] This is supported in the literature where the advantages of AF4 over SEC for the separation of macromolecules with high molar masses are demonstrated.[133]

The treatment of AF4 data involved creating a template using Origin software to calculate the weight-average molar masses (M_w) and dispersity ($Đ$) of molar mass distributions for each sample by employing the relevant equations.[143] The values obtained, however, were not congruent with literature values. BSA monomer was calculated to have a M_w of 38 000 g·mol^{-1} and the M_w of α-LA was calculated as 8 600 g·mol^{-1}. These values are approximately half of those published and are stipulated in Sections 1.6.4.2 (α-LA) and 1.6.4.3 (BSA). We have attributed this to the lack of recent

calibration of the MALS detector, and due to limited time and access to the instrument, calibration could not be performed to obtain a more accurate calibration constant following experimentation. The quantitative data obtained from the AF4 experiments is therefore merely used for comparison purposes between the two proteins. BSA had a higher dispersity ($Đ$=1.12) in terms of molar mass than α-LA ($Đ$=1.04). Although the values may not be accurate, the same trend is evident in SEC data (Table 2-1) and the results are supported by extensive literature.

FFF can be complementary to CE to assist in data interpretation in the same way SEC has been utilized in chapter 2. The direct coupling of FFF to CE for a 2D separation is currently being explored. Theoretically, FFF fractions can be collected and injected into the CE for the analysis of more homogenous fractions. Limited access to AF4 instrumentation and a restricted timeframe unfortunately meant the coupling of these techniques for this project was not possible.

4.4. Conclusions and future work

In this chapter, CN-PAGE confirmed the presence of multiple conformations in a sample of ADH. The presence of two separate bands corresponding to the dimeric form of the protein supports the that 2 types of dimers with slight differences in conformation can exist.[39] A two-dimensional separation of ADH, employing CN-PAGE in the first dimension and CE in the second, was performed as a proof of concept. A number of modifications can be made to the CE method, one example being electrokinetically injecting the sample which concentrates it.[144] In order for this method to be successful in its proposed application of purifying and/or homogenizing samples prior to analysis by CE, further work in assessing and overcoming the drawbacks faced in this study is required.

Pressure mobilization experiments revealed adsorption of catalase and DNAse I, resulting in the elimination of these two proteins for further analysis during this project. Adsorption is not an insurmountable drawback in CE and PM, and ways in

which it can be overcome have been described in chapters 2 and 5. When adsorption is minimal, TDA experiments can be conducted, providing insight into the weight-average hydrodynamic radius of the sample. Finally, AF4 of BSA and α-LA produced distributions of molar masses congruent with SEC results, indicating BSA is heterogenous with regards to molar mass, in comparison to α-LA which is more homogenous. AF4 experiments can be time consuming, depending on the nature of the sample and the expertise of the operator. Being a fairly well-established sizing technique, however, the potential to couple AF4 to CE for 2D separation is promising due to the many advantages AF4 has over techniques such as SEC and PAGE.

CHAPTER 5: Conclusions and Future Research

The overall aim of this work was to establish the method of free-solution capillary electrophoresis (CE-CC), to be used complementarily to techniques such as SEC, for the structural characterization of proteins. Currently, no single method is sufficient for the complete elucidation of the complex relationship that exists between protein structure and function.

5.1. Structural characterization of globular proteins

CE-CC was successfully used to obtain distributions of electrophoretic mobilities for three structurally well-characterized proteins: BSA, α-LA and ADH. The molar mass distributions for these proteins were also successfully obtained through SEC and AF4. For ADH, CN-PAGE revealed two populations of dimers which were not revealed by SEC. The heterogeneity of these proteins was quantified through the dispersion of their molar mass and electrophoretic mobility distributions. The dispersion of electrophoretic mobility distributions can be measured either by the dispersity value, where zero is used as a reference, or the standard deviation where the reference used is the weight average electrophoretic mobility. The choice of which measure of dispersion to use is ultimately dependent on the properties being studied. In protein chemistry and polymer science, dispersity is the preferred measure used for dispersion, while in separation science, the standard deviation may be more informative. These results revealed that molar mass distributions do not necessarily completely reflect the overall structural heterogeneity of a sample.

CE-CC can monitor the dissolution of lyophilized samples of BSA and insulin. The robustness of CE also enabled the analysis of the conformational equilibria, irrespective of the degree of dissolution. The results revealed that conformational equilibrium is not entirely dependent on the rate of dissolution and hence samples may require additional time to equilibrate. Upon refinement of the method, analysis of dissolution may be used to improve sample preparation, and/or as a supplementary means of characterization.

5.2. Further method development

CE-CC has been successfully implemented in the analysis of a number of natural (DNA, starch and chitosan) and synthetic (poly(acrylic acid)) polymers. CE of proteins in this work has been highly repeatable. To improve the overall analysis of proteins using CE however, further development of the method is required. This may involve tailoring the experimental parameters to obtain more accurate electrophoretic mobility distributions and/or coupling CE to other techniques for a more comprehensive characterization.

5.2.1. Tailoring separation parameters

For the majority of the proteins analyzed in this work, the results from PM experiments showed that adsorption on the capillary wall is negligible under the given conditions. Although adsorption is significantly reduced in CE due to the lack of a stationary phase, proteins may still interact with the inner surface of the capillary, hindering their migration. Several approaches can be employed to reduce these interactions, examples of which include incorporating additives to the BGE and/or altering the pH thereof. Another approach is to use coated capillaries. Agilent Technologies, for example, has commercialized capillaries coated with poly(vinyl alcohol) and a proprietary fluorocarbon polymer, both of which can be used for protein separations at neutral pH.

Altering the injection volume according to the concentration of the sample may be necessary to ensure adequate signal detection and at the same time avoid issues associated with overloading the capillary. This is important for improving the sensitivity of the detection. Another way to do so is through sample derivatization, however this may affect the mechanism of separation. It may also have an influence on the protein's conformational distribution and properties such as solubility.

Further work on formulating a specific standard that is separated in the same BGE as the sample could lead to a more reliable validation of the instrument and capillary as

well as increase the experimental throughput. Another aspect to consider could be implementing a different internal standard and mobility marker other than DMSO. From our results, it was suggested that DMSO could potentially be interacting with the proteins over extended periods of incubation and as such, may have had an impact on the results obtained in our preliminary analysis of dissolution and conformational equilibrium.

The application of CE to characterize the conformational heterogeneity of a range of globular proteins has been shown here to be highly reproducible across two instruments. Further investigations on the interlaboratory reproducibility of the method are of interest and will potentially continue in collaboration with the School of Purpan, Toulouse Institut National Polytechnique, France.

5.2.2. Furthering the utilization of other separation-based techniques to complement CE

In this work, results from SEC and AF4 experiments were primarily used in combination with literature to assist with the interpretation of the CE data. Additionally, CN-PAGE was able to provide insight into the conformations present under slightly varied conditions for a particular protein. Comparing the results from the different techniques gives a more holistic view of the structural heterogeneity of proteins.

For a more comprehensive exploration of the relationship between the properties characterized by these different techniques, however 2D coupling is of interest. Coupling, with CE-CC the second dimension, would allow separations first by size, then by electrophoretic mobility. The potential of CN-PAGE to be used in the first dimension was briefly explored in this thesis. Although these experiments were unsuccessful, a number of factors that could have led to this outcome are yet to be explored. Coupling CN-PAGE in the first dimension could also act as a means of sample purification prior to analysis by CE. Alternatively, AF4 could be employed in

the first dimension. This would be more advantageous as the retrieval of the analyte after the first separation is less tedious than in CN-PAGE, and the absence of a stationary phase eliminates the risk of shearing.

5.3. Applications and future outcomes

The CE method for protein analysis developed in this work may have significant implications namely in the biomedical, pharmaceutical, food and nutrition industries.

5.3.1. Utilizing the method for research and development purposes

The flexibility of the technique itself, as well as the low cost and comparably shorter analysis time makes CE exceptionally suited for research purposes. Advancements in biomedical research would come about from attaining a more complete characterization of proteins as this would assist in the elucidation of processes such as aggregation which has often been associated with disease.[120] A more comprehensive understanding of the roles of enzymes and other actively functional proteins could also offer insight into deficiencies and other disorders not necessarily related to aggregation. Consequently, the pharmaceutical industry could build upon these findings through subsequently improving or developing new drugs to treat or better manage those diseases and/or disorders. CE can also be used to investigate interactions between proteins and drugs, which could also lead to improvements in the design and effectivity of protein-based drug carriers. Similarly, analyzing the protein composition in food products could lead to enhanced quality and nutritional value.

5.3.2. Commercialization of the method

In addition to the contributions to research, the method could also be used commercially for routine analysis and quality assurance purposes. The robustness, high reproducibility and potential for it to be miniaturized are just some of the

features of CE that make it commercially appealing. In the biomedical industry for example, CE could be employed for diagnostic purposes. Additionally, evaluating and controlling the quality of products being manufactured is crucial in the pharmaceutical, food and nutrition industries.

CHAPTER 6: References

1. Nelson DL, Cox MM. Lehninger principles of biochemistry. 6th edition, North American: New York, N.Y.: W.H. Freeman and Company; 2013.
2. Mannige R. Dynamic new world: Refining our view of protein structure, function and evolution. Proteomes. 2014;2(1):128.
3. Ahram M, Springer DL. Large-scale proteomic analysis of membrane proteins. Expert review of proteomics. 2004;1:293-302.
4. Carver JA, Grosas AB, Ecroyd H, Quinlan RA. The functional roles of the unstructured n- and c-terminal regions in alphab-crystallin and other mammalian small heat-shock proteins. Cell Stress Chaperones. 2017;22(4):627-638.
5. Anaya Castro MA, Alric I, Brouillet F, Peydecastaing J, Fullana SG, Durrieu V. Soy protein microparticles for enhanced oral ibuprofen delivery: Preparation, characterization, and in vitro release evaluation. AAPS PharmSciTech. 2018;19(3):1124-1132.
6. De Oliveira MAL, Porto BLS, De A. Bastos C, Sabarense CM, Vaz FAS, Neves LNO, Duarte LM, Da S. Campos N, Chellini PR, Da Silva PHF, De Sousa RA, Marques R, Sato RT, Grazul RM, Lisboa TP, De O. Mendes T, Rios VC. Analysis of amino acids, proteins, carbohydrates and lipids in food by capillary electromigration methods: A review. Analytical methods. 2016;8(18):3649-3680.
7. Palopoli N, Monzon AM, Parisi G, Fornasari MS. Addressing the role of conformational diversity in protein structure prediction. PLoS ONE. 2016;11(5):e0154923.
8. Oldfield CJ, Yugong C, Cortese MS, Brown C, Uversky VN, Dunker AK. Comparing and combining predictors of mostly disordered proteins. Biochemistry. 2005;44(6):1989-2000.
9. Schad E, Tompa P, Hegyi H. The relationship between proteome size, structural disorder and organism complexity. Genome biology. 2011;12(12):Article ID R120.
10. Vaiana SM, Emanuele A, Palma-Vittorelli MB, Palma MU. Irreversible formation of intermediate bsa oligomers requires and induces conformational changes. Proteins: Structure, Function, and Bioinformatics. 2004;55(4):1053-1062.
11. Li M. Can we determine a protein structure quickly? Journal of Computer Science and Technology. 2010;25(1):95-106.
12. Boehr DD, Dyson HJ, Wright PE. An nmr perspective on enzyme dynamics. Chemical Reviews. 2006;106(8):3055-3079.
13. Kelly SM, Jess TJ, Price NC. How to study proteins by circular dichroism. BBA - Proteins and Proteomics. 2005;1751(2):119-139.

14. Gottarelli G, Lena S, Masiero S, Pieraccini S, Spada GP. The use of circular dichroism spectroscopy for studying the chiral molecular self-assembly: An overview. Chirality. 2008;20(3-4):471-485.

15. Shortt DW. Differential molecular weight distributions in high performance size exclusion chromatography. Journal of Liquid Chromatography. 1993;16(16):3371-3391.

16. Yamamichi S, Jinno Y, Haraya N, Oyoshi T, Tomitori H, Kashiwagi K, Yamanaka M. Separation of proteins using supramolecular gel electrophoresis. Chemical Communications. 2011;47(37):10344-10346.

17. Shevchenko A, Wilm M, Vorm O, Mann M. Mass spectrometric sequencing of proteins from silver-stained polyacry lamide gels. Analytical Chemistry. 1996;68(5):850.

18. Mahler HC, Friess W, Grauschopf U, Kiese S. Protein aggregation: Pathways, induction factors and analysis. Journal of Pharmaceutical Sciences. 2009;98:2909-2934.

19. Oliveira BM, Coorssen JR, Martins-de-Souza D. 2de: The phoenix of proteomics. Journal of Proteomics. 2014;104:140-150.

20. Chen JL, Morawetz H. Determination of molecular weight distribution of synthetic flexible-chain polyelectrolytes by polyacrylamide gel electrophoresis. Macromolecules. 1982;15(4):1185-1188.

21. Rabilloud T, Chevallet M, Luche S, Lelong C. Two-dimensional gel electrophoresis in proteomics: Past, present and future. Journal of Proteomics. 2010;73(11):2064-2077.

22. Sepsey A, Bacskay I, Felinger A. Polydispersity in size-exclusion chromatography: A stochastic approach. Journal of Chromatography A. 2014;1365:156-163.

23. Hong P, Koza S, Bouvier ESP. Size-exclusion chromatography for the analysis of protein biotherapeutics and their aggregates. Journal of Liquid Chromatography & Related Technologies. 2012;35(20):2923-2950.

24. Philo J. A critical review of methods for size characterization of non-particulate protein aggregates. Current Pharmaceutical Biotechnology. 2009;10:359-372.

25. Reschiglian P, Roda B, Zattoni A, Tanase M, Marassi V, Serani S. Hollow-fiber flow field-flow fractionation with multi-angle laser scattering detection for aggregation studies of therapeutic proteins. Analytical and Bioanalytical Chemistry. 2014;406(6):1619-1627.

26. Berek D. Size exclusion chromatography – a blessing and a curse of science and technology of synthetic polymers. Journal of Separation Science. 2010;33(3):315-335.

27. Bruessau RJ. Experiences with interlaboratory gpc experiments. Macromolecular Symposia. 1996;110(1):15-32.

28. Thevarajah JJ, Gaborieau M, Castignolles P. Separation and characterization of synthetic polyelectrolytes and polysaccharides with capillary electrophoresis. Advances in Chemistry. 2014;2014:Article ID 798503.

29. Oliver JD, Sutton AT, Karu N, Phillips M, Markham J, Peiris P, Hilder EF, Castignolles P. Simple and robust monitoring of ethanol fermentations by capillary electrophoresis. Biotechnology and Applied Biochemistry. 2015;62(3):329-342.

30. Oliver JD, Gaborieau M, Hilder EF, Castignolles P. Simple and robust determination of monosaccharides in plant fibers in complex mixtures by capillary electrophoresis and high performance liquid chromatography. Journal of chromatography A. 2013;1291:179-186.

31. Castignolles P, Gaborieau M, Hilder EF, Sprong E, Ferguson CJ, Gilbert RG. High-resolution separation of oligo(acrylic acid) by capillary zone electrophoresis. Macromolecular Rapid Communications. 2006;27(1):42-46.

32. Cottet H, Gareil P. From small charged molecules to oligomers: A semiempirical approach to the modeling of actual mobility in free solution. Electrophoresis. 2000;21(8):1493-1504.

33. Grossman PD, Colburn JC, Lauer HH, Nielsen RG, Riggin RM, Sittampalam GS, Rickard EC. Application of free-solution capillary electrophoresis to the analytical scale separation of proteins and peptides. Analytical Chemistry. 1989;61(11):1186-1194.

34. Geiger M, Hogerton AL, Bowser MT. Capillary electrophoresis. Analytical Chemistry. 2012;84(2):577.

35. Harstad RK, Johnson AC, Weisenberger MM, Bowser MT. Capillary electrophoresis. Analytical Chemistry. 2016;88(1):299-319.

36. Rollet M, Gle D, Phan TNT, Guillaneuf Y, Bertin D, Gigmes D. Characterization of functional poly(ethylene oxide)s and their corresponding polystyrene block copolymers by liquid chromatography under critical conditions in organic solvents. Macromolecules. 2012;45(17):7171-7178.

37. Thevarajah JJ, Sutton AT, Maniego AR, Whitty EG, Harrisson S, Cottet H, Castignolles P, Gaborieau M. Quantifying the heterogeneity of chemical structures in complex charged polymers through the dispersity of their distributions of electrophoretic mobilities or of compositions. Analytical Chemistry. 2016;88(3):1674-1681.

38. Williams SKR, Caldwell K. Field-flow fractionation in biopolymer analysis. Wien New York: Springer; 2012.

39. Raj SB, Ramaswamy S, Plapp BV. Yeast alcohol dehydrogenase structure and catalysis. Biochemistry. 2014;53(36):5791-5803.

40. Plapp BV, Charlier HA, Ramaswamy S. Mechanistic implications from structures of yeast alcohol dehydrogenase complexed with coenzyme and an alcohol. Archives of Biochemistry and Biophysics. 2016;591:35-42.

41. Hayes JE, Velick SF. Yeast alcohol dehydrogenase: Molecular weight, coenzyme binding, and reaction equilibria. The Journal of biological chemistry. 1954;207(1):225.

42. Permyakov EA, Berliner LJ. A-lactalbumin: Structure and function. FEBS Letters. 2000;473(3):269-274.

43. Slangen CJ, Visser S. Use of mass spectrometry to rapidly characterize the heterogeneity of bovine alpha-lactalbumin. Journal of Agricultural and Food Chemistry. 1999;47(11):4549.

44. Pike ACW, Brew K, Acharya KR. Crystal structures of guinea-pig, goat and bovine α-lactalbumin highlight the enhanced conformational flexibility of regions that are significant for its action in lactose synthase. Structure. 1996;4(6):691-703.

45. Chrysina ED, Brew K, Acharya KR. Crystal structures of apo- and holo-bovine alpha-lactalbumin at 2. 2-a resolution reveal an effect of calcium on inter-lobe interactions. Journal of Biological Chemistry. 2000;275(47):37021-37029.

46. Wijesinha-Bettoni R, Gao C, Jenkins JA, Mackie AR, Wilde PJ, Mills ENC, Smith LJ. Heat treatment of bovine alpha-lactalbumin results in partially folded, disulfide bond shuffled states with enhanced surface activity. Biochemistry. 2007;46(34):9774.

47. Majorek KA, Porebski PJ, Dayal A, Zimmerman MD, Jablonska K, Stewart AJ, Chruszcz M, Minor W. Structural and immunologic characterization of bovine, horse, and rabbit serum albumins. Molecular Immunology. 2012;52(3-4):174-182.

48. Barone G, Capasso S, DelVecchio P, DeSena C, Fessas D, Giancola C, Graziano G, Tramonti P. Thermal denaturation of bovine serum albumin and its oligomers and derivatives ph dependence. Journal of Thermal Analysis. 1995;45(6):1255-1264.

49. Borzova VA, Markossian KA, Chebotareva NA, Kleymenov SY, Poliansky NB, Muranov KO, Stein-Margolina VA, Shubin VV, Markov DI, Kurganov BI. Kinetics of thermal denaturation and aggregation of bovine serum albumin. PLoS ONE. 2016;11(4):Article ID 0153495.

50. Babcock JJ, Brancaleon L. Bovine serum albumin oligomers in the e- and b-forms at low protein concentration and ionic strength. International Journal of Biological Macromolecules. 2013;53:42-53.

51. Malamud D, Drysdale JW. Isoelectric points of proteins: A table. Analytical Biochemistry. 1978;86(2):620-647.

52. Frankær CG, Knudsen MV, Norén K, Nazarenko E, Ståhl K, Harris P. The structures of t6, t3r3 and r6 bovine insulin: Combining x-ray diffraction and absorption spectroscopy. Acta Crystallographica Section D. 2012;68(10):1259-1271.

53. Bryant C, Spencer D, Miller A, Bakaysa DL, McCune KS, Maple, Sr., Pekar A, Brems D. Acid stabilization of insulin. Biochemistry. 1993;32(32):8075-8082.

54. Xu Y, Yan Y, Seeman D, Sun L, Dubin PL. Multimerization and aggregation of native-state insulin: Effect of zinc. Langmuir. 2012;28(1):579-586.

55. Conway-Jacobs A, Lewin LM. Isoelectric focusing in acrylamide gels: Use of amphoteric dyes as internal markers for determination of isoelectric points. Analytical Biochemistry. 1971;43(2):394-400.

56. Prakash K, Prajapati S, Ahmad A, Jain SK, Bhakuni V. Unique oligomeric intermediates of bovine liver catalase. Protein Science. 2002;11(1):46-57.

57. Calderón I, Arenas F, Pérez J, Fuentes D, Araya M, Saavedra C, Tantaleán J, Pichuantes S, Youderian P, Vásquez C. Catalases are nad(p)h-dependent tellurite reductases. PLoS One. 2006;1(1):Article ID e70.

58. Schroeder WA, Shelton JR, Shelton JB, Olson BM. Some amino acid sequences in bovine-liver catalase. BBA - Enzymological Subjects. 1964;89(1):47-65.

59. Samejima T, Kamata M, Shibata K. Dissociation of bovine liver catalase at low ph. Journal of biochemistry. 1962;51:181-187.

60. Oefner C, Suck D. Crystallographic refinement and structure of dnase i at 2 r… resolution. Journal of Molecular Biology. 1986;192(3):605-632.

61. Suck D, Oefner C, Kabsch W. Three-dimensional structure of bovine pancreatic dnase i at 2.5 (angstrom) resolution. EMBO Journal. 1984;3(10):2423-2430.

62. Kim HS, Liao T-H. Isoelectric focusing of multiple forms of dnase in thin layers of polyacrylamide gel and detection of enzymatic activity with a zymogram method following separation. Analytical Biochemistry. 1982;119(1):96-101.

63. Biemann K. Contributions of mass spectrometry to peptide and protein structure. Biological Mass Spectrometry. 1988;16(1-12):99-111.

64. Benesch JLP, Robinson CV. Mass spectrometry of macromolecular assemblies: Preservation and dissociation. Current Opinion in Structural Biology. 2006;16(2):245-251.

65. Wüthrich K. Protein structure determination in solution by nmr spectroscopy. Journal of Biological Chemistry. 1990;265(36):22059-22062.

66. Castellani F, van Rossum B, Diehl A, Schubert M, Rehbein K, Oschkinat H. Structure of a protein determined by solid-state magic-angle-spinning nmr spectroscopy. Nature. 2002;420(6911):98-102.

67. Drenth J. Principles of protein x-ray crystallography: Springer New York; 2007.

68. Frank J. Single-particle imaging of macromolecules by cryo-electron microscopy. Annual review of biophysics and biomolecular structure. 2002;31(1):303-319.

69. Laue TM. Analytical ultracentrifugation. Current Protocols in Protein Science. 2001:7.5. 1-7.5. 9.

70. Wen J, Arakawa T, Philo JS. Size-exclusion chromatography with on-line light-scattering, absorbance, and refractive index detectors for studying proteins and their interactions. Analytical Biochemistry. 1996;240(2):155-166.

71. Haslbeck M, Franzmann T, Weinfurtner D, Buchner J. Some like it hot: The structure and function of small heat-shock proteins. Nature structural & molecular biology. 2005;12(10):842-846.

72. Holt C, Carver JA, Ecroyd H, Thorn DC. Invited review: Caseins and the casein micelle: Their biological functions, structures, and behavior in foods. Journal of Dairy Science. 2013;96(10):6127-6146.

73. Baumann M, Meri S. Techniques for studying protein heterogeneity and post-translational modifications. Expert review of proteomics. 2004;1(2):207-217.

74. Svergun DI, Koch MHJ, Timmins PA, May RP. Small angle x-ray and neutron scattering from solutions of biological macromolecules: OUP Oxford; 2013.

75. Jehle S, Vollmar BS, Bardiaux B, Dove KK, Rajagopal P, Gonen T, Oschkinat H, Klevit RE. N-terminal domain of αb-crystallin provides a conformational switch for multimerization and structural heterogeneity. Proceedings of the National Academy of Sciences. 2011;108(16):6409-6414.

76. Hochberg GK, Benesch JL. Dynamical structure of αb-crystallin. Progress in biophysics and molecular biology. 2014;115(1):11-20.

77. Gilbert RG, Hess M, Jenkins AD, Jones RG, Kratochvil R, Stepto RFT. Dispersity in polymer science. Pure and Applied Chemistry. 2009;81(4):779-779.

78. Harrisson S. The downside of dispersity: Why the standard deviation is a better measure of dispersion in precision polymerization. Polymer Chemistry. 2018;9(12):1366-1370.

79. Koppel DE. Analysis of macromolecular polydispersity in intensity correlation spectroscopy: The method of cumulants. The Journal of Chemical Physics. 1972;57(11):4814-4820.

80. Frisken BJ. Revisiting the method of cumulants for the analysis of dynamic light-scattering data. Applied Optics. 2001;40(24):4087-4091.

81. Grossman PD, Colburn JC, Lauer HH, Nielsen RG, Riggin RM, Sittampalam G, Rickard EC. Application of free-solution capillary electrophoresis to the analytical scale separation of proteins and peptides. Analytical Chemistry. 1989;61(11):1186-1194.

82. Simpson DC, Smith RD. Combining capillary electrophoresis with mass spectrometry for applications in proteomics. Electrophoresis. 2005;26(7-8):1291-1305.

83. Fonslow BR, Yates Iii JR. Capillary electrophoresis applied to proteomic analysis. Journal of Separation Science. 2009;32(8):1175-1188.

84. Cottet H, Gareil P, Theodoly O, Williams CE. A semi-empirical approach to the modeling of the electrophoretic mobility in free solution: Application to polystyrenesulfonates of various sulfonation rates. Electrophoresis. 2000;21:3529-3540.

85. Karim MR, Shinagawa S, Takagi T. Electrophoretic mobilities of the complexes between sodium dodecyl sulfate and various peptides or proteins determined by free solution electrophoresis using coated capillaries. Electrophoresis. 1994;15(1):1141-1146.

86. Stellwagen NC, Gelfi C, Righetti PG. The free solution mobility of DNA. Biopolymers. 1997;42(6):687-703.

87. Gaborieau M, Causon TJ, Guillaneuf Y, Hilder EF, Castignolles P. Molecular weight and tacticity of oligoacrylates by capillary electrophoresis–mass spectrometry. Australian Journal of Chemistry. 2010;63(8):1219-1226.

88. Castignolles P, Gaborieau M, Hilder EF, Sprong E, Ferguson CJ, Gilbert RG. High-resolution separation of oligo(acrylic acid) by capillary zone electrophoresis. Macromolecular Rapid Communications. 2006;27:42-46.

89. Chamieh J, Martin M, Cottet H. Quantitative analysis in capillary electrophoresis: Transformation of raw electropherograms into continuous distributions. Analytical Chemistry. 2015;87(2):1050-1057.

90. Thevarajah JJ, Van Leeuwen MP, Cottet H, Castignolles P, Gaborieau M. Determination of the distributions of degrees of acetylation of chitosan. International Journal of Biological Macromolecules. 2017;95:40-48.

91. Mnatsakanyan M, Thevarajah JJ, Roi RS, Lauto A, Gaborieau M, Castignolles P. Separation of chitosan by degree of acetylation using simple free solution capillary electrophoresis. Analytical and Bioanalytical Chemistry. 2013;405(21):6873-6877.

92. Maniego AR, Ang D, Guillaneuf Y, Lefay C, Gigmes D, Aldrich-Wright JR, Gaborieau M, Castignolles P. Separation of poly (acrylic acid) salts according to topology using capillary electrophoresis in the critical conditions. Analytical and Bioanalytical Chemistry. 2013;405(28):9009-9020.

93. Toutounji MR, Van Leeuwen MP, Oliver JD, Shrestha AK, Castignolles P, Gaborieau M. Quantification of sugars in breakfast cereals using capillary electrophoresis. Carbohydrate Research. 2015;408:134-141.

94. Weinberger R. Practical capillary electrophoresis. 2nd ed. San diego, CA: Academic Press; 2000.

95. Sutton AT, Arrua RD, Gaborieau M, Castignolles P, Hilder EF. Characterization of oligo(acrylic acid)s and their block co-oligomers. Analytica Chimica Acta. 2018;1032:163-177.

96. Cherney LT, Krylov SN. Slow-equilibration approximation in studying kinetics of protein adsorption on capillary walls. Analyst. 2015;140(8):2797-2803.

97. Delaittre G, Save M, Gaborieau M, Castignolles P, Rieger J, Charleux B. Synthesis by nitroxide-mediated aqueous dispersion polymerization, characterization, and physical core-crosslinking of ph- and thermoresponsive dynamic diblock copolymer micelles. Polymer Chemistry. 2012;3(6):1526-1538.

98. Gaborieau M, Castignolles P. Size-exclusion chromatography (sec) of branched polymers and polysaccharides. Analytical and Bioanalytical Chemistry. 2011;399(4):1413-1423.

99. Berek D. Size exclusion chromatography - a blessing and a curse of science and technology of synthetic polymers. Journal of Separation Science. 2010;33(3):315-335.

100. Thevarajah JJ, Bulanadi JC, Wagner M, Gaborieau M, Castignolles P. Towards a less biased dissolution of chitosan. Analytica Chimica Acta. 2016;935:258-268.

101. Taylor DL, Ferris CJ, Maniego AR, Castignolles P, in het Panhuis M, Gaborieau M. Characterization of gellan gum by capillary electrophoresis. Australian Journal of Chemistry. 2012;65:1156-1164.

102. Cherney LT, Petrov AP, Krylov SN. One-dimensional approach to study kinetics of reversible binding of protein on capillary walls. Analytical Chemistry. 2015;87(2):1219-1225.

103. Oliver JD, Rosser AA, Fellows CM, Guillaneuf Y, Clement JL, Gaborieau M, Castignolles P. Understanding and improving direct uv detection of monosaccharides and disaccharides in free solution capillary electrophoresis. Analytica Chimica Acta. 2014;809:183-193.

104. de Frutos M, Cifuentes A, Diez-Masa JC, Camafeita E, Mendez E. Multiple peaks in hplc of proteins: Bovine serum albumin eluted in a reversed-phase system. Hrc-Journal of High Resolution Chromatography. 1998;21(1):18-24.

105. Yohannes G, Wiedmer SK, Elomaa M, Jussila M, Aseyev V, Riekkola ML. Thermal aggregation of bovine serum albumin studied by asymmetrical flow field-flow fractionation. Analytica Chimica Acta. 2010;675(2):191-198.

106. Magonet E, Hayen P, Delforge D, Delaive E, Remacle J. Importance of the structural zinc atom for the stability of yeast alcohol-dehydrogenase. Biochemical Journal. 1992;287:361-365.

107. Gong X, Xiong X, Qi L, Fang X. Investigating the structural transitions of proteins during dissolution by mass spectrometry. Talanta. 2017;164:418-426.

108. Kramer RM, Shende VR, Motl N, Pace CN, Scholtz JM. Toward a molecular understanding of protein solubility: Increased negative surface charge correlates with increased solubility. Biophysical Journal. 2012;102(8):1907-1915.

109. Pau B, Martin B. Structural biology: Proteins in dynamic equilibrium. Nature. 2010;468(7327):1046-1048.

110. Yang S, Blachowicz L, Makowski L, Roux B. Multidomain assembled states of hck tyrosine kinase in solution. Proceedings of the National Academy of Sciences. 2010;107(36):15757-15762.

111. Levi V, González Flecha FL. Reversible fast-dimerization of bovine serum albumin detected by fluorescence resonance energy transfer. BBA - Proteins and Proteomics. 2002;1599(1):141-148.

112. Yongbo Z, Erik RPZ. The 70-kda heat shock protein chaperone nucleotide-binding domain in solution unveiled as a molecular machine that can reorient its functional subdomains. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(28):10272-10277.

113. Roy I, Gupta MN. Freeze-drying of proteins: Some emerging concerns. Biotechnology and Applied Biochemistry. 2004;39(2):165-177.

114. Baker LJ, Hansen AMF, Bhaskara Rao P, Bryan WP. Effects of the presence of water on lysozyme conformation. Biopolymers. 1983;22(7):1637-1640.

115. Costantino H, Langer R, Klibanov A. Moisture-induced aggregation of lyophilized insulin. Pharm Res. 1994;11(1):21-29.

116. Schmitz S, Dona AC, Castignolles P, Gilbert RG, Gaborieau M. Assessment of the extent of starch dissolution in dimethyl sulfoxide by 1h nmr spectroscopy. Macromolecular Bioscience. 2009;9(5):506-514.

117. Maniego AR, Sutton AT, Gaborieau M, Castignolles P. Assessment of the branching quantification in poly(acrylic acid): Is it as easy as it seems? Macromolecules. 2017;50:9032-9041.

118. Wong TM, Carey CM, Lin SHC. Rapid characterization of soy protein and hydrolysates by capillary electrophoresis. Journal of Chromatography A. 1994;680(2):413-417.

119. Harvey D. Modern analytical chemistry: Boston: McGraw-Hill Companies, Inc.; 2000.

120. Sabella S, Quaglia M, Lanni C, Racchi M, Govoni S, Caccialanza G, Calligaro A, Bellotti V, De Lorenzi E. Capillary electrophoresis studies on the aggregation process of β-amyloid 1-42 and 1-40 peptides. Electrophoresis. 2004;25(18-19):3186-3194.

121. Brange J, Andersen L, Laursen ED, Meyn G, Rasmussen E. Toward understanding insulin fibrillation. New York1997. p. 517-525.

122. Fink AL. Protein aggregation: Folding aggregates, inclusion bodies and amyloid. Folding and Design. 1998;3(1):R9-R23.

123. Smisek DL, Hoagland DA. Agarose gel electrophoresis of high molecular weight, synthetic polyelectrolytes. Macromolecules. 1989;22(5):2270-2277.

124. Wittig I, Schägger H. Features and applications of blue-native and clear-native electrophoresis. Proteomics. 2008;8(19):3974-3990.

125. Chamieh J, Oukacine F, Cottet H. Taylor dispersion analysis with two detection points on a commercial capillary electrophoresis apparatus. Journal of Chromatography A. 2012;1235:174-177.

126. Hawe A, Hulse W, Jiskoot W, Forbes R. Taylor dispersion analysis compared to dynamic light scattering for the size analysis of therapeutic peptides and proteins and their aggregates. Pharmaceutical Research. 2011;28(9):2302-2310.

127. Le Saux T, Cottet H. Size-based characterization by the coupling of capillary electrophoresis to taylor dispersion analysis. Analytical Chemistry. 2008;80(5):1829-1832.

128. Chamieh J, Cottet H. Comparison of single and double detection points taylor dispersion analysis for monodisperse and polydisperse samples. Journal of Chromatography A. 2012;1241:123-127.

129. Ibrahim A, Meyrueix RM, Pouliquen G, Chan Y, Cottet H. Size and charge characterization of polymeric drug delivery systems by taylor dispersion analysis and capillary electrophoresis. Analytical and Bioanalytical Chemistry. 2013;405(16):5369-5379.

130. Chetwynd AJ, Guggenheim EJ, Briffa SM, Thorn JA, Lynch I, Valsami-Jones E. Current application of capillary electrophoresis in nanomaterial characterisation and its potential to characterise the protein and small molecule corona. Nanomaterials. 2018;8(2):DOI 10.3390/nano8020099.

131. Chmelik J. Applications of field-flow fractionation in proteomics: Presence and future. Proteomics. 2007;7(16):2719-2728.

132. Giddings JC, Yang FJ, Myers MN. Theoretical and experimental characterization of flow field-flow fractionation. Analytical Chemistry. 1976;48(8):1126-1132.

133. Otte T, Brull R, Macko T, Pasch H, Klein T. Optimisation of ambient and high temperature asymmetric flow field-flow fractionation with dual/multi-angle light scattering and infrared/refractive index detection. Journal of Chromatography A. 2010;1217(5):722-730.

134. Williams SKR, Benincasa M-A, Smith WC, Oliver JD. Field-flow fractionation in analysis of polymers and rubbers. Encyclopedia of analytical chemistry: John Wiley & Sons, Ltd; 2006.

135. Flory PJ. Principles of polymer chemistry. Ithica, N.y.: Cornell University Press; 1953.

136. Schaertl W. Light scattering from polymer solutions and nanoparticle dispersions. Berlin, Heidelberg: Springer-Verlag; 2007.

137. Ihara M, Matsuura N, Yamashita A. High-resolution native-page for membrane proteins capable of fluorescence detection and hydrodynamic state evaluation. Analytical Biochemistry. 2011;412(2):217-223.

138. Jörnvall H. The primary structure of yeast alcohol dehydrogenase. European Journal of Biochemistry. 1977;72(3):425-442.

139. Tal M, Silberstein A, Nusser E. Why does coomassie brilliant blue r interact differently with different proteins? A partial answer. The Journal of biological chemistry. 1985;260(18):9976.

140. de Jong S, Krylov SN. Pressure-based approach for the analysis of protein adsorption in capillary electrophoresis. Analytical Chemistry. 2012;84(1):453-458.

141. Zhao H, Brown PH, Schuck P. On the distribution of protein refractive index increments. Biophysical Journal. 2011;100(9):2309-2317.

142. Tumolo T, Angnes L, Baptista MS. Determination of the refractive index increment (dn/ dc) of molecule and macromolecule solutions by surface plasmon resonance. Analytical Biochemistry. 2004;333(2):273-279.

143. Gaborieau M, Gilbert RG, Gray-Weale A, Hernandez JM, Castignolles P. Theory of multiple detection size exclusion chromatography of complex branched polymers. Macromolecular Theory and Simulations. 2007;16:13-28.

144. Breadmore MC, Sanger-van de Griend CE, Majors RE. In-capillary sample concentration in ce. LC-GC North America. 2014;32(3):174.

145. Perlmann GE, Longsworth LG. The specific refractive increment of some purified proteins. Journal of the American Chemical Society. 1948;70(8):2719.

146. Stuting HH, Krull IS. Complete on-line determination of biopolymer molecular weight via high-performance liquid chromatography coupled to low-angle laser light scattering, ultraviolet, and differential refractive index detection. Analytical Chemistry. 1990;62(19):2107.

147. Masayuki N, Yoh S. Light scattering studies on the thermal denaturation of bovine serum albumin. Bulletin of the Chemical Society of Japan. 1973;46(3):791-797.

148. Wagoner DB, Christman RF. Molar masses and radii of humic substances measured by light scattering. Acta hydrochimica et hydrobiologica. 1998;26(3):191-195.

149. Barer R, Joseph S. Refractometry of living cells: Part i. Basic principles. Journal of Cell Science. 1954;95(4):399-423.

Appendix

Table A-1 Information of protein samples studied in the thesis provided by the supplier. pI is the isoelectric points given by the supplier. The extinction coefficient of a 1% solution ($E^{1\%}$) was given by the supplier at the following wavelengths: [A] 280 nm, [B] 279 nm, [C] 276 nm, [D] 278 nm.

Protein	CAS number	Supplier	Catalog number	$E^{1\%}$	pI_{supp}	$Molar\ mass_{supp}$ ($g \cdot mol^{-1}$)
ADH	9031-72-5	Sigma-Aldrich	A7011	14.6[A]	5.4-5.8	141 000-151 000
α-LA	9051-29-0	Sigma-Aldrich	L6010	20.1[A]	4.5	14 000
BSA	9048-46-8	Sigma-Aldrich	A7517	6.67[B]	4.7-5.3	66 000
Catalase	9001-05-2	Sigma-Aldrich	C1345	36.5[C]	5.4	250 000
DNAse I	9003-98-9	Sigma-Aldrich	D4263	11.1[A]	4.8-5.2	30 000
Insulin	11070-73-8	Sigma-Aldrich	I5500	10.6[D]	5.3	5 733

Table A-2 Summary of values related to separation quality by CE. The average adjusted R^2 values relate to the quality of Gaussian fit on the elugrams obtained by PM. Relative recovery is calculated as percent difference in the peak areas between CE and PM experiments for each protein.

Protein	Instrument	Injection volume (nL)	Sample Concentration ($g \cdot L^{-1}$)	Average Adjusted R^2	Relative recovery (%)
ADH	Sciex MDQ (n=2)	11.44	0.93	0.995	15.3
	Agilent 7100 (n=7)	13.42	1.29	0.990	38.7
				0.978	34.1
α-LA	Sciex MDQ (n=2)	11.44	1.00	0.999	35.2
	Agilent 7100 (n=7)	13.42	1.00	0.992	28.8
				0.999	21.3
BSA	Sciex MDQ (n=2)	11.44	0.99	0.980	22.8
	Agilent 7100 (n=2)	13.42	1.00	0.998	9.30

Table A-3 Summary of characterization of two proteins by CE using different injection volumes. Dispersion of the samples are shown as D(W(μ), 1,0) and D(W(μ), 2,0) in terms of electrophoretic mobility. The error was estimated as the standard deviation for n=2.

Protein	Injection pressure and duration	Injection volume (nL)	Dispersity ($D(W(\mu),1,0)$)	Dispersity ($D(W(\mu),2,0)$)	Average $D(W(\mu),1,0$	Weight average mobility (μ_w)	Average μ_w
α-LA	13.79 mbar for 8.7 s	~ 4.58	1.014	1.018	1.013 ± 0.002	1.42E-08	1.40E-08 ± 2.46E-10
			1.012	1.014		1.38E-08	
	6.89 mbar for 5 s	~ 1.31	1.017	1.023	1.023 ± 0.007	1.99E-08	1.80E-08 ± 2.68E-09
			1.028	1.035		1.61E-08	
BSA	13.79 mbar for 8.7 s	~ 4.58	1.004	1.004	1.004 ± 0.001	2.18E-08	2.08E-08 ± 1.44E-09
			1.005	1.006		1.98E-08	
	6.89 mbar for 5 s	~ 1.31	1.013	1.014	1.011307 ± 0.002	2.30E-08	2.27E-08 ± 3.92E-10
			1.011	1.013		2.28E-08	
			1.009	1.011		2.22E-08	

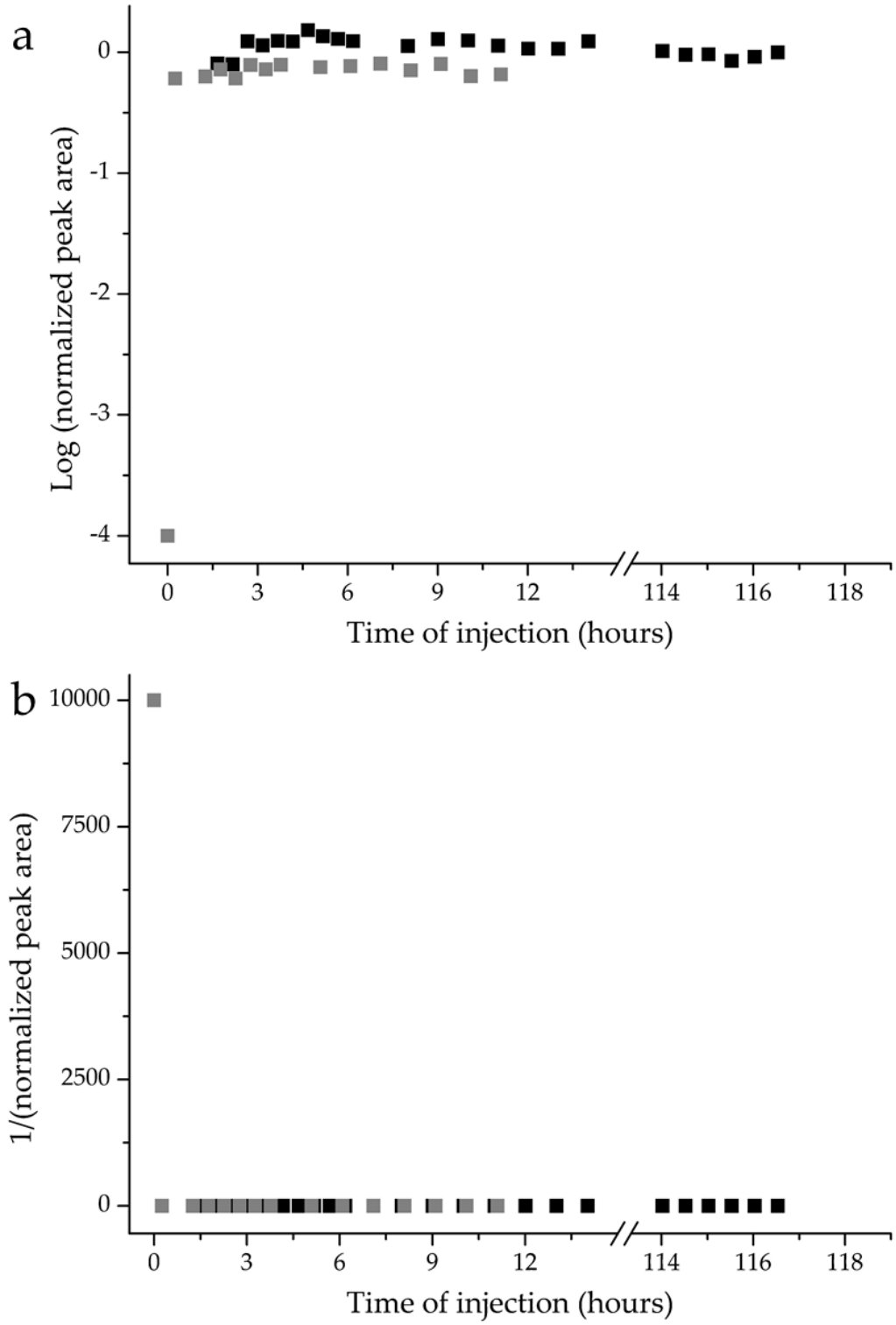

Figure A-1 Plots created to determine whether the dissolution of BSA follows a) first-order, or b) second order kinetics in PB5, pH 7.4.

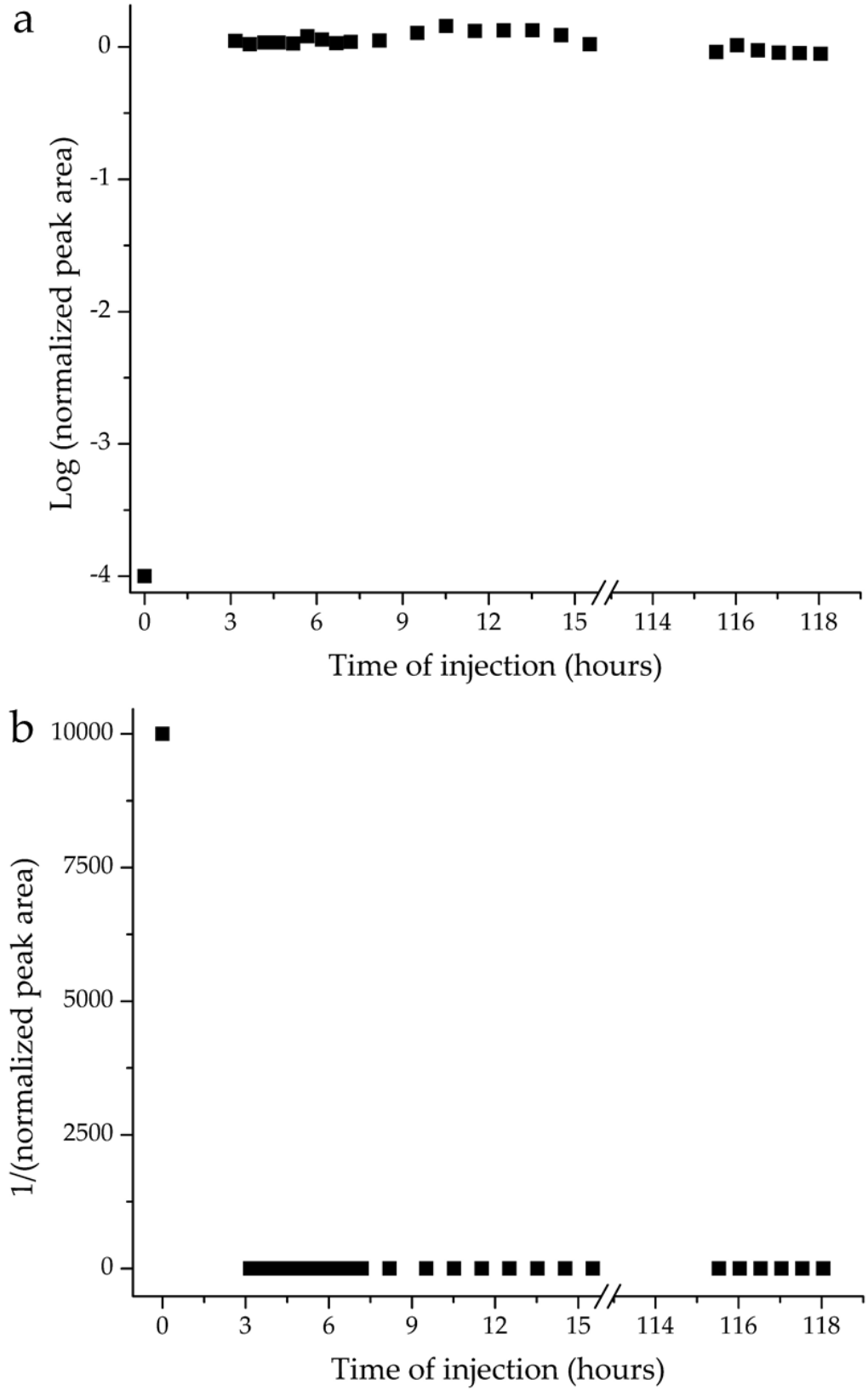

Figure A-2 Plots created to determine whether the dissolution of insulin follows a) first-order, or b) second order kinetics in PB5, pH 7.4.

Table A-4 Signal-to-noise ratios of oligo(sodium acrylate) for experiments conducted 5 days apart

Date	Signal-to-noise ratio
12/09/2018	2.67
17/09/2018	2.50

Table A-5 $D(W(\mu),1,0)$, $D\sigma$, μ_w and FWHM for replicate CE experiments under analogous conditions to those in Chapter 2

Protein	Dispersity ($D(W(\mu),1,0)$)	Standard Deviation $D\sigma$	Weight average mobility (μ_w)	Normalized FWHM
Insulin	1.012	1.51E-9	1.41E-8	-
	1.012	1.53E-9	1.41E-8	-
	1.025	2.35E-9	1.49E-8	-
	1.023	2.25E-9	1.49E-8	-
Average	1.018	1.91E-09	1.45E-08	
Standard deviation	6.98E-3	4.52E-10	4.62E-10	
RSD %	0.68	23	3.19	
BSA	1.005	1.23E-9	1.71E-8	-
	1.005	1.22E-9	1.71E-8	-
	1.008	1.77E-9	1.86E-8	-
	1.005	1.68E-9	1.71E-8	-
	1.005	1.37E-9	1.99E-8	1.810
	1.005	1.54E-9	1.98E-8	1.786
	1.013	2.71E-9	2.30E-8	1.627
	1.012	2.62E-9	2.28E-8	1.534
	1.010	2.31E-9	2.22E-8	-
	1.013	2.37E-9	1.97E-8	-
	1.015	2.58E-9	1.97E-8	-
	1.016	2.63E-9	1.97E-8	-
	1.016	2.63E-9	1.97E-8	-
	1.015	2.64E-9	2.01E-8	-
Average	1.010	2.093E-09	1.975E-08	1.689
Standard deviation	4.58E-3	5.88E-10	1.94E-9	0.139
RSD %	0.45	28	9.8	8.2

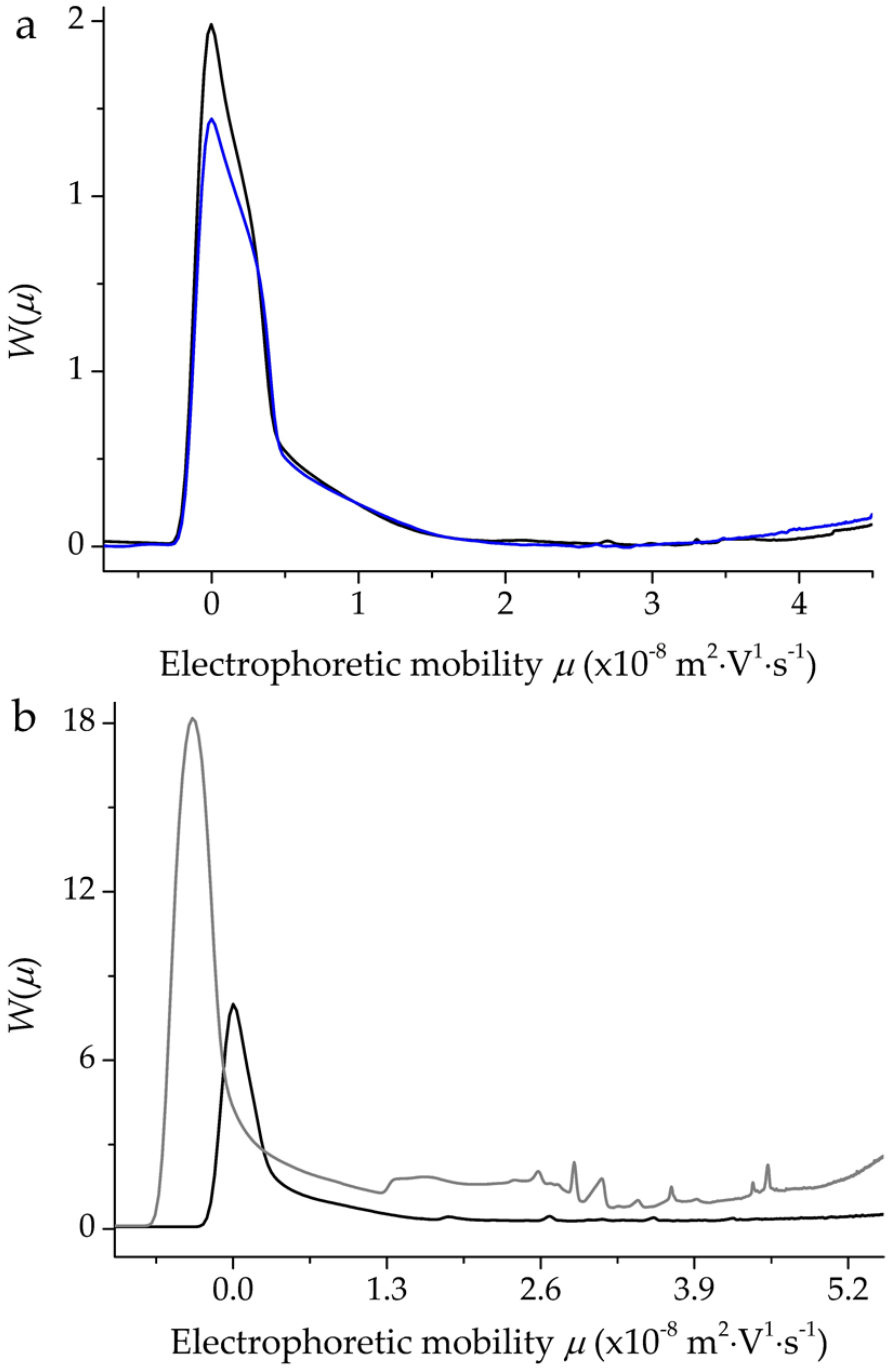

Figure A-4 Electropherograms of ADH samples eluted from CN-PAGE gel for: a) fractions containing no CBB (light and dark blue correspond to replicates), and b) fractions containing CBB (black and gray correspond to separation times of 10 min and 30 min, respectively). The peak at 0 $m^2 \cdot V^{-1} \cdot s^{-1}$ corresponds to DMSO.

Table A-6 Refractive index increments of BSA as published by various sources

Solvent	Wave-length (nm)	Temp (°C)	Concen-tration	d*n*/d*c*	Remarks	Source
Water	578	0	60.59 g·L^{-1}	0.190	pH 5.05	Sigma Product Information sheet
Water	578	20	Not specified	0.187	Not specified	Publication[145]
Water	578	25	Not specified	0.187	Not specified	Publication[145]
Water + 0.1 M NaCl	Not specified	Not specified	52.19 g·L^{-1}	1.93	pH 5.35	Publication[145]
Water + 0.5 M NaCl	Not specified	Not specified	46.34 g·L^{-1}	1.95	pH 5.31	Publication[145]
PB 8.5 + 150 mM NaCl	488	20	1-2.5 g·L^{-1}	0.186	pH 7.4	Publication[105]
Phosphate buffer (aq)	633	Sub-ambient	3-5 g·L^{-1}	0.168	Buffer: (PB50 + 150 mM NaCl), pH 7.2	Refractive index Data-book (2000), reference 32[146]

0.05 M Phosphate buffer	436	25	0.25% BSA in soln.	0.194	pH 7.2	Publication [147]
0.1 M Phosphate buffer (aq.)	690	Not specified	5-15 $g \cdot L^{-1}$	0.174	Buffer: 0.05 M Na2HPO4 + 0.05 M KH_2PO_4, pH 5.8	Refractive index Data-book, reference 199[148]
Water + 0.1 M NaCl	546	25	Not specified	0.185	pH 5.2	Refractometry of living cells Handbook[149]